



mdw - University of Music and Performing Arts Vienna
Austrian Academy of Sciences (ÖAW)

21.-24. Mai 2026

Eleventh International Conference on Technologies
for Music Notation and Representation
Edited by Elisabeth Schimana



TENOR

INTERNATIONAL CONFERENCE
ON TECHNOLOGIES FOR MUSIC
NOTATION & REPRESENTATION



The Eleventh International Conference on Technologies for Music Notation and Representation
Vienna, Austria 21–24 May 2026

Proceedings published by mdw – University of Music and Performing Arts Vienna
ISBN 978-0-6481592-2-3

This conference was funded in whole or in part by the Austrian Science Fund (FWF)
[grant DOI 10.55776/AR824].

<https://soundscore.org/TENOR26/>
<https://soundscore.org/>
<https://tenor-conference.org>

These proceedings are licensed under the terms of the Creative Commons Attribution 4.0 International CC BY
License. <http://creativecommons.org/licenses/by/4.0>

Preface

Researchers, composers, scientists, technologists and performers presented their work at the International Conference on Technologies for Music Notation and Representation TENOR 2026. This conference in Vienna had a focus on innovative advancements on audio scores: musical notation, representation technologies, methods of composition, interpretation, psychoacoustic questions and aesthetic analysis connected to the concept of mimesis. The conference is part of the research project Sound as Score funded in whole or in part by the Austrian Science Fund (FWF) and hosted by the University of Music and Performing Arts Vienna (mdw) and the Austrian Academy of Sciences (ÖAW), Vienna.

Out of more than 100 submissions, 30 papers, 9 workshops, 4 posters and 18 sonic works were presented at TENOR 26 Vienna.

We made an effort to make the conference accessible to researchers and artists who are not affiliated with a university by awarding grants.

Special thanks go to Florian Bogner for bringing the compositions to life at the Future Art Lab Klangtheater and the ensembles *airborne extended*, *Black Page Orchestra*, *ANNEA* and the *Soundscore Ensemble*, formed specifically for the conference and composed of students from the mdw, under the direction of Jaime Wolfson.

We thank all the researchers and artists responding to our call, and all the authors presented in this proceedings, and Johannes Kretz and Theresia Dlouhy-Staber from the ARC (mdw - Artistsic Research Center) and Piotr Majdak from the ARI (ÖAW - Acoustics Research Institute) for their organizational support.

And we would especially like to thank the reviewers for their expert and excellent evaluation of the submissions and Johannes Meissl, Vice Rector of the mdw, for his outstanding support of the TENOR 2026 conference.

Elisabeth Schimana
Program chair

Committee

Program chair:

Elisabeth Schimana (mdw – Artistic Research Center ARC)

Scientific committee:

Alex Hofmann (mdw - Department of Music Acoustics IWK)

Susanne Kogler (University of Graz KFU / Department of Arts and Musicology)

Piotr Majdak (Austrian Academy of Sciences ÖAW / Acoustics Research Institute ARI)

Katharina Pollack (Austrian Academy of Sciences ÖAW / Acoustics Research Institute ARI)

Artistic chair:

Thomas Grill (mdw – Artistic Research Center ARC)

Organisation:

Barbara Maria Neu (mdw – Artistic Research Center ARC)

Ernst Reitermaier (mdw – Artistic Research Center ARC)

Nicolas Müller-Lorenz (University of Graz KFU / Department of Arts and Musicology)

TENOR International Steering Committee

Rama Gottfried

Georg Hajdu

Cat Hope (General Chair)

Craig Vear

ZHdK, Zurich – Switzerland

Hochschule für Musik und Theater Hamburg – Germany

Monash University, Melbourne – Australia

Nottingham University – UK

Reviewers

Agostini Andrea
Akkermann Miriam
Anatrini Alessandro
Artacho Adrian
Bača Trevor
Basica Constantin
Beers Jean
Bresson Jean
Brüggemann Tanja
Caulet David
Celestini Federico
Chuang Se-Lien
Cox Jessie
D'Errico Lucia
Doettlinger Marco
Dufek Hannes
Esling Philippe
Fields Ken
Forciniti Isabella
Giavitto Jean-Louis
Gstättner-Heckel Maria

Haffter Christoph
Haselboeck Lukas
Hassfurther Sophie
Hausch Cora
Hildebrand Nolan
Hofmann Alex
Josefson Mathias
Kim-Boyle David
Kocher Philippe
Kogler Susanne
Kretz Johannes
Križić Ivar Roban
Lach Juan
Lechner Patrik
Majdak Piotr
Markovic Nenad
Mendivil Julio
Mendoza Mejia Enrique
Moroz Solomiya
Noisternig Markus
Olarte Luis Alejandro

Plessas Peter
Pollack Katharina
Saunders James
Schröder Gesine
Sdraulig Charlie
Sepperer Markus
Sioros George
Sköld Mattias
Tian Fu
Tomás Enrique
Tremblay Pierre Alexandre
Turowski Paul
Tveit Anders
Urbanek Nikolaus
Venus Peter
Vickery Lindsay
Weixler Andreas
Whalley Ian
Winter Michael

Table of Contents

An Exploration of Audio-Score Techniques and Practices	1
<i>Lindsay Vickery</i>	
Listening Guides as Compositional Devices: Textual and Formal Strategies in <i>Les Dispartions</i>	13
<i>Sébastien Roux</i>	
Living Audio Scores: Designing Listening-Based Notation in Nada Sfera Playground	19
<i>Ainolnaim Azizol</i>	
Trying to Keep Up: Managing Cognitive Load in Audio Scores	25
<i>James Saunders</i>	
Experiment Design for Psychoacoustic Analyses of Audio-Score Compositions	31
<i>Katharina Pollack, Piotr Majdak, Barbara Maria Neu</i>	
Composing the Ears of the Machine: Audio Scores, Selective Listening, and Concatenative AI Improvisation.....	36
<i>Ming Yang</i>	
Generative Audioscores as Engines of Social Choreography in “Villanelles de Voyelles II”	45
<i>Sandeep Bhagwati, Kasey Pocius</i>	
Interpreting Audio Scores: Experiences and Insights	53
<i>Barbara Maria Neu</i>	
Playing the Room: Tapeplayer and the Improvised Audio Score	58
<i>Jonas Sjøvaag</i>	
Information Embedment in Live-Electronics: Scores, Setups and Performer Knowledge	64
<i>Alex Hofmann, Sonja Stojak</i>	
Towards a Notation of Sceno-electroacoustic Works.....	72
<i>Rémy Jannelle, Nicolas Bernier</i>	
Iterative Score for Digital Sound Composition	77
<i>Peter Plessas</i>	
The “Ragion Tropica” Tactics: Inhabiting the Vibration, Audio Cartography, Sound Routes, and Other Useless Hacking Tools	83
<i>Giuseppe Silvi</i>	
Mass-spring Lattice Physical Modeling Synthesis Using a Quantum Algorithm.....	94
<i>Xavier Davenport</i>	
Cricket Tools: Real-Time Audio Scoring of Quantum State Evolution Across Networked Mobile Devices	101
<i>Pedro González-Fernández, Thomas Noll, Reiko Yamada</i>	
Beyond Notation: Li-Yue, From Centralized Spectacle to a Community Model.....	107
<i>Eric Chi-Wei Lin</i>	
Mutable Scores: Embracing Non-Finality in Notation	114
<i>Alexandre Sasset-Blouin</i>	
(re-)Defining the Poetic Text Score.....	121
<i>Craig Pedersen</i>	
Sound as Score in Architectural Plans: a Path-based Notation Workflow Using Ambisonics + 360° Video.....	133
<i>Nicolaj Kirisits</i>	

Sonic Scores Mimetic Behaviour: Two Case Studies	140
<i>Eric Maestri</i>	
Songs of Love and Death: The Act of Sound as Notation	146
<i>Gil Dori</i>	
Interactive Scores: Tasks and Decisions in Improvisation Systems and Open Works	150
<i>Per Anders Nilsson, Palle Dahlstedt</i>	
Visual Scores as Listening-Mediated Notation: A Practic-Based Case Study	157
<i>Grgur Savić</i>	
Prescriptive Notation as Performative Interface	163
<i>Djordje Marković</i>	
Playing the Action, Hearing the Result: Prescriptive Notation and Bassoon Technique in Aaron Cassidy's <i>25 April '05</i>	170
<i>James Aylward</i>	
F Operator: A Proposal for a Generative Notation for Frequency Factors in Music	178
<i>Markus Lepper, Baltasar Trancón Y Widemann</i>	
A Prosodic Layer for Suprasegmental Features Spanning Rhythm Trees, Motivated by Carnatic Rhythmical Structures	190
<i>Daniel Miller</i>	
From Rhythm to Representation: BTS's "Idol" as a Multimodal Audio Score	202
<i>Veronika Mattová, Michal Schwarz</i>	
Individualizing Graphic Notation through Botanical Metaphors: A Practice-Based Study of <i>Botánica</i>	208
<i>Jingsong Teng</i>	
SIGN/e: Symbol-based Interface for Graphic Notation Edition for Ableton Live	215
<i>Pierre-Luc Lecours, Nicolas Bernier, Evan Montpellier, Phillipe-Aubert Gauthier</i>	

AN EXPLORATION OF AUDIO-SCORE TECHNIQUES AND PRACTICES

Lindsay Vickery

Edith Cowan University
l.vickery@ecu.edu.au

ABSTRACT

This paper discusses the emergence of the Audio Score as a distinct compositional practice in which auditory communication, rather than visual notation, is the primary medium through which compositional information is conveyed to performers. Drawing on recent theoretical writing by Bhagwati, Sdraulig and Lortie, Schimana, and d’Heudières, the study situates Audio Scores within the evolution of technological, aesthetic, and ideological developments.

The discussion proposes a framework distinguishing three core roles of auditory material in Audio Scores—orientation, emulation, and instruction—and seeks to clarify the boundaries between Audio Scores and adjacent practices. The analysis foregrounds performer experience as a critical site of evidence through detailed case studies of works by d’Heudières, Schimana, and the author, all performed by the Australian ensemble *GreyWing*.

1. INTRODUCTION

This paper examines the emergence of the Audio Score, an overarching term capturing a range of practices, techniques and aesthetics that have only recently been identified and codified in the work of Bell [1], Bhagwati [2], Sdraulig and Lortie [3], Schimana [4] and d’Heudières [5]. The paper examines functional roles and criteria for Audio Score classification, the relationship between technological advances and the emergence of the Audio Score and examines four works performed by the Ensemble *GreyWing* as recent case studies. Where possible I have connected my thoughts to practical experiences of performing Audio Scores (and what might be termed proto-Audio Scores).

The core definition of the term Audio Score discussed by these authors, focuses on the shift towards the primacy of “auditory communication to convey (...) composer-defined aspects (of a creative work) to the performers”, [2] rather than visual communication through notation: Common Practice (CPN), graphic, text or other means. This shift represented a significant reconfiguration of where authority resides in the communication of musical ideas.

It should be noted that it is implied that this communication takes place through technologically-based

auditory communication – that technological mediation is not incidental but constitutive to the medium. The approach was made possible as the result of the development of pertinent audio technologies and is distinct from transmission of music ideas orally or “by ear” (presumably the oldest and still most common form of communication of musical information). The defining criterion of an Audio Score is not simply the presence of audio per se, but the primacy of auditory communication in determining performative action.

Bhagwati [2] and Sdraulig and Lortie [3] have both created categorisations of the types of recorded information and affordances found in Audio-Scores. In relation to their core role as auditory communication, these categories can be summarised as employing verbal or sonic information with only three types of function, each with different implications in regard to the balance of composer/performer authority: sound objects (the performer is delegated interpretive authority to imitate, interact with and/or exemplify sounds), orientation (cues orienting the performer in time implying a minimal shift in authority) and instructions (the composer exerts procedural authority byverbally directing actions to be taken by the performer) (see Table 1).

Verbal	Sonic
Orientation: ie “One”, “Two measures before A”, “Part Two”.	Orientation: ie metronome click, bell.
Instructions: ie “begin playing”, “imitate”, “describe”.	Sound Objects: ie sonic exemplars to be imitated, interacted with or used as exemplars.

Table 1. A summary of the functions of Verbal and Sonic cues or prompts in Audio-Scores

In terms of the works discussed in the five foundational papers mentioned above, it is important to stress that although Audio Scores may include orientative information, it is generally only the presence of either verbal instructions, sound objects or both that exemplifies the primacy of auditory communication at the core of the definition. Works that principally employ audio to provide temporal orientation for another medium, for example a film score click track, lie outside of the genre definition, because in most cases they preserve notational authority rather than assigning primary authority to sound itself.

Like many examples of emerging terminology, the term “Audio Score” is still somewhat in flux, with both examples of its use to describe different practices, and examples of different terminology being used for the same or similar practices. The first appearance of the term “Audio Score” in the literature that is clearly related to the

	verbal		sonic		delivery		performance		additional score			sounding		structure	
	orientation	instruction	orientation	emulation	audible	inaudible	rehearsed	extempore	CPN	graphic	text	audio	environment	fixed	indeterminate
Robert Erickson - <i>Pacific Sirens</i> (1968)				x	x		x		x	x		x		x	
Gavin Bryars - <i>1, 2, 1-2-3-4</i> (1971)				x		x	x								x
Alvin Lucier - <i>Carbon Copies</i> (1989)				x	x	x	x	x						x	x
Lindsay Vickery - <i>Night Fragments</i> (2011)	x		x	x		x	x		x			x		x	
Lionel Marchetti - <i>Inland Lake</i> (2019)				x	x		x					x		x	
Cat Hope - <i>Tough it Out</i> (2014)		x				x		x		x					x
Louis d'Heudieres - <i>Laughter Studies 3</i> (2016)		x		x		x	x	x						x	
Lindsay Vickery - <i>Un Paso al Abismo</i> (2018)				x		x	x							x	
Elisabeth Schimana - <i>Landscapes</i> (2021)				x		x	x							x	
Lindsay Vickery - <i>little grey wing</i> (2022)				x		x	x			x				x	

Table 2. Audio Score classification criteria for the works to be discussed in this paper.

post-2018 definition, seems to occur in an interview with performance artist and technologist Laurie Anderson about Robert Wilson's production *Alceste* (1986). [6]

Anderson uses the term to describe an audio recording created from actors' speech, and "sound effects, such as birds and wind" that was used to provide cues to assist in coordinating live performers in the stage production. Its existence draws attention to the parallel lineage of audio-based performer direction in performance art and theatre, that will be briefly discussed below.

In music, a similar early example (but not labelled or conceptualised as an Audio Score) was Stockhausen - *Kontakte* (1958–60) [7], in which the performers use both "synchronized clocks and their ears to synchronize with the tape". [8] Anderson's use of the term was isolated, and underlines the problem that the coordination of live performers with fixed media that first emerged in the motion picture industry in the 1930s, was addressed by numerous solutions that sought to solve a technological issue rather than an aesthetic one. This may have resulted in a lack of documentation of their development and the circumstance that only some of these solutions clearly lead toward the development of the Audio Score as defined by Bhagwati et al.

In addition to the Verbal and Sonic possibilities outlined above, there are a number of complimentary criteria that, while not defining the Audio Score, shape how its authority is experienced by performers and audiences.

- **Audible - Inaudible delivery/transmission to the audience.** For many examples of the Audio Score, transmission is "individual" via headphones, mirroring the conventional practice of the textual score in performance. As with notated scores, individual transmission creates the situation in which the audience are unaware of the internal details and scaffolding of the work, whereas a shared audio score can also function as a component of the musical fabric of the sounding work. In some cases, separate audio is also presented to the audience in addition to the performer's inaudible Audio Score.
- **Rehearsed/Extempore Performance.** One of the key advantages of communicating musical ideas through sound itself is the ability to efficiently model or create complex timbres, temporal information and continuous parametrical changes for performers to emulate. However, the situations resulting from performers responding to unfamiliar audio or

rehearsed audio presented in unfamiliar sequences are also exploited in some works.

- **Audio Score Alone/Additional Score or Audio Materials.** Although auditory communication is core to the Audio score, perhaps the most contentious boundary case is whether the addition of other kinds of scored materials place a work beyond the definition of the term by undermining the primacy of audio.

As is often the case, actual examples of creative works occupy every continuum of these practices. This paper will discuss my own observations in regard to a number of Audio Score works that I have had first-hand experience performing, composing or both. These works are classified in Table 2. according to the criteria for Audio Score classification outlined above.

2. PRECURSOR TECHNOLOGIES, AESTHETICS AND WORKS

I have elsewhere argued that the appropriate technological and ideological conditions were necessary for the emergence of works exploring the combination of acoustic instruments with field recordings. [9] Similarly, the emergence of the Audio Score and its various sub-genres was dependent on the establishment of specific technological capabilities and ideological shifts and these evolved well before the Audio Score concept was named or codified in detail. Once these conditions were established, a range of broadly similar approaches and practices emerged independently among unrelated projects in art music, performance art, dance and theatre. The following discussion plots three principal threads.

2.1. Orientation

Bell [1] and Bhagwati [2] have pointed to the click track, invented to synchronise orchestral scores with motion pictures, as an important precursor technology for the orientative function of the Audio Score. Foundational technologies necessary for the development of the click-track included the electronic transmission of audio signals begun with the telegraph and the telephone; recording; and headphones (For detailed discussion of their development see [10, 11, 12, 13, 14]). The desire to coordinate live performance with fixed media in art music, thus retaining centralized compositional authority, became evident with

the emergence of the early works for instruments and fixed media (tape) such as Bruno Maderna - *Musica su due dimensioni* (1952), [15] Edgard Varèse - *Déserts* (1954), [16] and Luigi Nono - *La Fabbrica Illuminata* (1964). [17] The exact date of adoption of click track into art music is difficult to trace, as early examples may have emerged to solve performative technological coordination problems (rather than compositional and aesthetic ones), and as such were not often discussed in the literature or recorded by composers as an integrated component of their works. An exception is Ghent's creation of an analogue system using "a magnetic tape recording on which signals to the individual performers (had) been pre-recorded at different pitch levels' allowing them to perform with "complete independence as to tempo, meter, and positioning of the beat, and yet be precisely coordinated in time" and "as a means of synchronising electronic tape music with live performers". [18] This electronic apparatus was ambitiously used to coordinate musicians performing a polytemporal score, dancers, fixed media and lighting in his work *Phosphores* (1967).

It is often claimed (including by myself) [19] that Xenakis coordinated the six percussionists in *Persephassa* (1969) via click track, but this appears not to have been the case, for the premiere performance at least (which used stopwatches). [20] The confusion may be the result of subsequent performances that used a click track as a coordination aid. The same is possibly true of Elliot Carter - *Second String Quartet* (1959), [21] for the reasons suggested above.

What is clear is that, at latest, the technique of click track coordination in art music began to be established shortly after the foundation of IRCAM (1977-) for example the instruments and fixed media work *Arcus* (1978) by York Höller. [22]

These technologies, while foundational to the development of the Audio Score solved a temporal coordination problem, (also useful to the Audio Score), but not yet a communicative and interpretive problem.

2.2. Emulation

Audio recording and reproduction technologies were also gingerly, if more conspicuously, adopted into art music, beginning with gramophone Respighi's *Pini Di Roma* (1925) [23, 24], Vertov's *Collages* [25, 26], Toch and Hindemith's *Gramophonmusik* (1930) [27], Cage's turntable, radio and electronics works [28, 29], Schaeffer and Henry's *Music Concrète* works [30] and other directions [31, 32, 33, 34] and laid the groundwork both technologically and ideologically for the emergence of a medium in which the textual score was diminished in importance, and the "sound itself" was to become the predominant medium for communicating the score.

The shift towards emulating the complexities of sound itself can be considered a proto-Audio Score strategy. It partially redistributes authority in that the performer's perception, and instrumental limitations mediate the result, as evident in Karlheinz Stockhausen - *Spiral* (1968) in which the performer is instructed to listen "quietly to broadcasts on the short-wave receiver (while searching one can join in quietly with instrument/voice), until one

has found something one wants to start with" and then to "join in as much and as precisely as they can". [35] Similarly in *Pacific Sirens* (1968), Robert Erickson directs "improvising" performer's to "listen into" "the spectral complexes of the environmental noise". [36, 37] Both works included a scored component: for *Spiral* this is a "plus-minus" meta-composition score - instructing how the performer is to transform the found sound that they are emulating, for *Pacific Sirens* the performers are given a quasi-spectrographic guide score showing pitch, duration and morphology of prominent frequencies and are instructed to "blend" with and "protrude" [36] from the recording. (Erickson does not discuss the technique used to create the "quasi-spectrographic" score in his *Sound Structure in Music* [38], only the band-filter processing of the ocean recording, but it is possible he created the score using the filter output or the "Kay Sona-Graph" a device with similar graphic output that was commonly used for speech and animal call analysis in the late 1960s).

In both *Spiral* and *Pacific Sirens*, the audience hears the electronic audio component in addition to the live performers, however three years later, in what is clearly an early example of core Audio Score practice, Gavin Bryars - *1, 2, 1-2-3-4* (1971) asks the performers to interact privately with an individually recorded copy of "the same piece or type" of "familiar music". "Instrumentalists/vocalists, each wearing headphones connected to a portable cassette machine. Each performer hears only the music in his headphones, music which contains "parts" for his instrument or voice, and he plays, along with the cassette, his own instrumental part. [39] (In the influential 1975 *Obscure Records* release they used "identical material" [40])

The process for realising the work is documented in a Text Score (it is framed as a text score in a collection called "Verbal Anthology" [39] and not merely "instructions"), but in concert, the performers interact exclusively with individual recorded audio. The process invariably results in a deconstruction of the source audio in which the players are inconsistently de-synchronised from one another due to the indeterminate start position and playback speed of the music on each cassette machine, and the fact that the players synchronise with their personal audio and not with one another. Here the listener isolation, so intrinsic to headphone listening, is exploited as structural material.

Synchronisation, or in this case precise desynchronisation is an important Audio Score affordance exploited by Bryars. It becomes a key element of many later Audio Score works including the Nancarrow-like structures of Joshua Fried - *Headset Sextet* (1995) [41] and Benedict Mason - *felt|ebb|thus|brink|here|array|telling* (2004). [1] In his notes, Bryars also foreshadows another important parameter: whether the performers have rehearsed their part or are responding in realtime. The rehearsed/extemporary condition and experience of the performer is also crucial in later theatrical works such as Ant Hampton's solo show *Blake* (1999) in which the audience watch an actor who receives instructions on stage in real time, so that the audience understand that he is "discovering everything at the same time as they are" [42] or Cat Hope - *Tough it Out* (2014) in which the audience

observes the spectacle of real-time instructions disrupting musicians as they attempt to read as score.

Another early example of the Audio Score is Alvin Lucier - *Carbon Copies* (1989), [43] which expands the emulation of acoustic environment found in his (*Hartford*) *Memory Space* (1970) [44], but in this case using a recording of an “indoor or outdoor environment” rather than relying on the performer’s memory. As in *Memory Space* the performers are instructed to imitate the sounds (this time on a self-made field recording) “as exactly as possible, without embellishment”. *Carbon Copies* also anticipates a crucial parameter of Audio Score practice by designing a process in which every configuration of the environmental sound is examined: in *Section I* both the audience and performers listen to the audio through loudspeakers; in *Section II* the audience continues to listen through loudspeakers while the performers listen through headphones and begin to emulate the recording; *Section III* the audience only hear the performers emulate the recording in their headphones; and *Section IV* the performers too, remove their headphones and continue “playing, either from memory or by freely imagining the sounds continuing”.

These early “emulative” works lay the groundwork for Lionel Marchetti’s *Partition Concrète* works for Decibel New Music Ensemble and others – in which an audible audio component created, from “sounds themselves” using the techniques of *Musique Concrète*, also functions as an “auditory score” for instrumentalists. [45]

2.3 Experiential Evidence

I participated in performances of *Pacific Sirens* in 2018 with *GreyWing* Ensemble and *1, 2, 1-2-3-4* in 1986 with *Tonight’s Ensemble* and *Carbon Copies* in 2010 with *Decibel*, and can comment that there are a clear experiential differences distinguishing them. In *Pacific Sirens*, there is an interplay between the audio component, the score and the performers always informing your choices of pitch, duration, timbre and dynamic – the performers are also listening to the complete work, creating a hybrid distribution of authority between the score, environment and performer agency. *1, 2, 1-2-3-4* on the other hand invites the performer to focus on maintaining synchronisation with the audio in their headphone and not to be influenced by the slightly varied temporal orientation of the other performers, restraining the performer’s authority via individual synchronisation. *Carbon Copies* brings a heightened awareness of the audience experience even when listening and performing from personal headphones, and shifts authority toward the memory and imagination of the performer. The Bryars and Lucier works both reward detailed practice: Bryars to assist maintaining focus and Lucier because of the potential complexity of finding appropriate techniques to emulate the environmental recording, while Erickson’s work permits a more exploratory and improvisational approach, perhaps due to the dominance of the ambient recorded component, but also because of the presence of a guide score.

2.3. Instructions

The rapid evolution of audio technology from 1970 - the advent of low cost, portable stereo and multi-track tape recorders in the late 1970s, such as the Sony Walkman [46] and TEAC 144 Portastudio, [47] the advent of microcomputing and sequencing in the 1980s, [48, 49] Digital Audio editing [50, 51, 52] and object oriented interactive software [53] in the 1990s and a general shift toward “private listening”, [54, 55, 56] introduced further possibilities for synchronisation and complexity and created the conditions for a second wave of Audio Scores. The point at which either orientative or instructive verbal cues were added to the metronome style click track is also not recorded, although my own first experience was in a 1986 performance of Steve Reich’s *Tehillim* (1981) by Nova Ensemble, at this time only possible because an ensemble member directed a multitrack recording studio. This performance in Perth, Australia suggests that these techniques were established quite rapidly following the developments at IRCAM or already existed and/or were transferred from the film industry.

As an unambiguous example, my own song cycle *Night Fragments* (2011) performed by Caitlin Cassidy with Decibel New Music Ensemble included verbal and electronic temporal cues, poly-temporal sections and pitch cues for the singer to assist her maintaining pitch within a chromatic cluster. I would argue that even what might be termed an “elaborate click track” such as this, despite being integral to performance of the work, does not constitute an Audio Score.

Bell gives the closer example of Aaron Cassidy - *I, purples, spat blood, laugh of beautiful lips* (2012), [1] which despite its complex score is dependent on a pitch received via headphones and varied in each performance. The potential for alteration of this work through its reliance on indeterminately generated pitch brings it closer, however I would still argue that auditory communication is not the primary source of its identity.

To Bell and Bhagwati’s discussion of the click track we should add the affordances of “Interruptible Foldback” (IFB) [57] systems that evolved as part of mid-20th-century broadcast studio engineering practices. These are audio cueing networks that provide mix-minus program feeds and director interruptions to on-air talent, [58] that generate the opportunity to create dynamic, interruptive authoritarian hierarchies.

The possibility of providing realtime delivery of verbal instructions with performers has also been explored in Cat Hope - *Tough it Out* (2014). It was composed for a concert of music celebrating Australia’s first female prime minister Julia Gillard, who faced a sadly predictable chorus of “noise and interference” in the course of her tenure. In *Tough it Out*, Hope uses this frustrating situation as a structural device. Performers are asked to read “any Cat Hope score, or (...) the score provided” [59] – a graphical score comprising prime ministerial popularity graphs – but are continually interrupted by verbal instructions received individually over headphones. The theatrical and conceptual potential of the situation is presided over by a dedicated electronics operator who chooses when and which of 17 messages to send to each

player throughout the work. Some of the instructions are more typically “musical” – “play like a whisper - super soft, almost not there”, while others can lead to more dramatic situations “copy one person next to you, whatever they are doing, as exactly as possible. Don’t hide it”, [59] a juvenile bullying tactic employed by Gillard’s political opposition. In addition to it commentary on the challenging tenure of Australia’s first female Prime Minister, the work serves as critique of authority, hierarchy, and control that is embedded in audio cueing systems and exemplifies the interdependence of technological and ideological developments.

3. CASE STUDIES

This section examines four Audio Score works relatively recently performed and/or recorded by the Western Australian New Music ensemble *GreyWing* (flute - Kirsten Smith, clarinet - Lindsay Vickery, Electric Guitar - Jameson Feakes and Harp - Catherine Ashley). The discussion analyses the techniques employed in each work and the structural ideas that are created in light of the experiential aspect of performing the works. The works are Louis d’Heudières - *Laughter Studies 3* (2016) [60] and Lindsay Vickery - *Un Paso al Abismo* (2018) [61] both part of a series of work exploring the use of text in music, Elisabeth Schimana – *Landscapes* (2021) [62] performed as part of a series of environmental and “instrument a field recording works” *Banksias under Snow* and *Little (Grey) Wing* (2021) [63] a hybrid Audio Score plus Scrolling score performed as part of the *Outcome Unknown* series. These works, while not exploring every corner of Audio Score practice exemplify some important nuanced differences in practice, including: the explicit use of instructions and addition of a reference score; the challenges of vocal in comparison to instrumental emulation; and a range of distinctive structural features that are exemplified in these works.

3.1. Louis d’Heudières’ - *Laughter Studies 3* (2016)

Louis d’Heudières *Laughter Studies 3* (2016) for three performers, uses synchronised individual verbal

instructions and sound objects to be emulated entirely vocally by the performers either by describing or imitating the sounds they hear. For the most part, this performative situation aligns with what Sdraulig and Lortie describe as “Reactive Emphasis”, in which “instructive/ performative prompts” elicit a “real-time improvisatory response to sound [that] performers cannot hear or infer ahead”. [2] However, d’Heudières does suggest that some sections be rehearsed as they benefit “from interpretation beyond merely reacting to the sounds”. [60]

Two of d’Heudières’ detailed instructions in particular, rely heavily on real-time improvisation. The instruction “describe” requires the performer to speak continuously for the duration of the sound - “like a stream of consciousness” - while “describe with one word only” demands that performers “go with the first word that pops into your head” [60], deliberately creating situations of temporal uncertainty and cognitive pressure. In the former case, performers must sustain improvisatory verbal description without knowing when the sound will end; in the latter, the task resembles a psychological word-association test, heightening performative stress. This pressure becomes materially significant in the final section (see Figure 4. below), when all three performers receive identical verbal and sonic prompts and their responses can be compared by the audience.

The setup exemplifies what Saunders terms a “self-evident process”—a work whose “processes [are] made self-evident [to the audience] through actions” [64]. The performers’ attentive listening through headphones and their audible reactions are sufficiently transparent that audiences can plausibly infer shared stimuli and compare divergent responses, often with comic effect. It creates the strongest example of the four works of a distributed and indeterminate authority in which the performers cognition becomes a compositional variable and the balance of composer/performer authority becomes clearly observable by the audience.

Across the seven *Laughter Studies* (2015–19), d’Heudières categorises sonic materials into four groups: “everyday sounds, ambient environments, music, and emotive sounds” [65]. *Laughter Studies 3* focuses

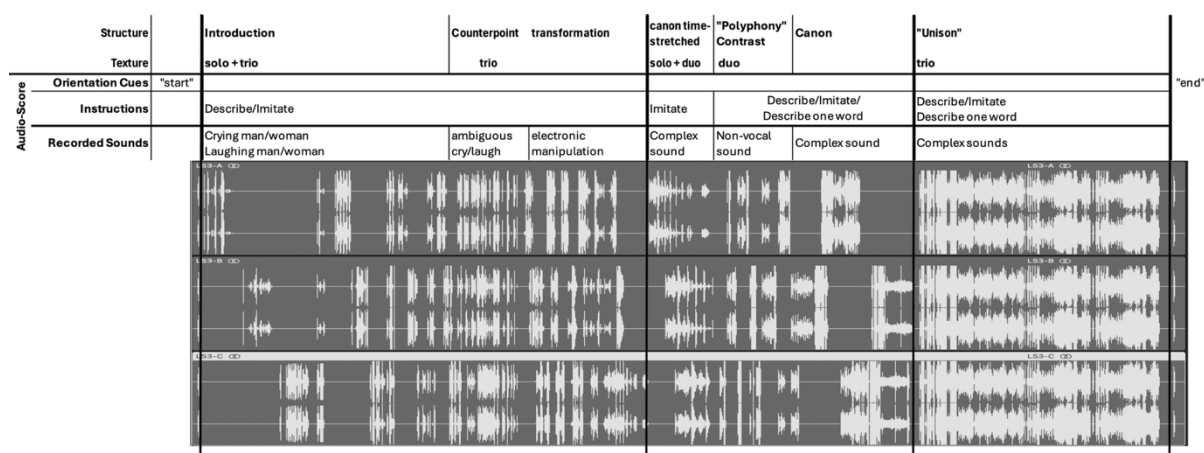


Figure 1. Louis d’Heudières - *Laughter Studies 3* (2016), Structural Analysis of the Audio-Scores for Performers 1, 2 and 3.



Figure 2. Louis d'Heudieres - *Laughter Studies 3* (2016), Counterpoint section: showing the interplay of the samples (numbered) and the instructions ("Imitate" and "Describe").

exclusively on everyday and emotive sounds. The emotive materials include recordings of crying and laughing at varying intensities, alongside a third, ambiguous state d'Heudieres terms "*Cry-Singing*" [60]: positioned between laughter and crying, this material complicates real-time description and imitation. The everyday sounds - ranging in duration from 2 to 37 seconds - are sometimes subjected to processing such as pitch-shifting and time-stretching, and include sources that resist straightforward vocal imitation or verbal description (for example often processed, footsteps on twigs, drumbeats, a double bass, clattering cutlery or the sound of a cork being removed). A structural analysis of the work (see Figure 1.) can be made based on: the structural role of each section; the number of simultaneous voices (Texture); and the type of information conveyed in the Audio-Score, plotted against a soundwave visualisation of the Audio-Score "parts" for each performer.

The work opens with something of a typical "introduction of forces" - each performer is heard in turn demonstrating the two principal instructions "describe" and "imitate". In the "counterpoint section, the recorded sounds become more ambiguous and difficult to describe, often resulting in the performers creating "non-lexical interjections ('uh', 'er', 'um'), awkward pauses in the middle of sentences and frequent involuntary hand gestures in the performers' interpretations". [65] A total of eight samples are used in a variety of combinations and in simultaneous or staggered orientations between the performers. The final part of this section settles into simultaneous imitations of contrasting samples (see Figure 2. above).

Laughter Studies 3 also features a more explicitly polyphonic section, in which a source sample (sometimes stretched or deconstructed) is heard in successive voices d'Heudieres states that "the acts of vocalised mimicry in *Laughter Studies* are specific, personal, and nonlinguistic.

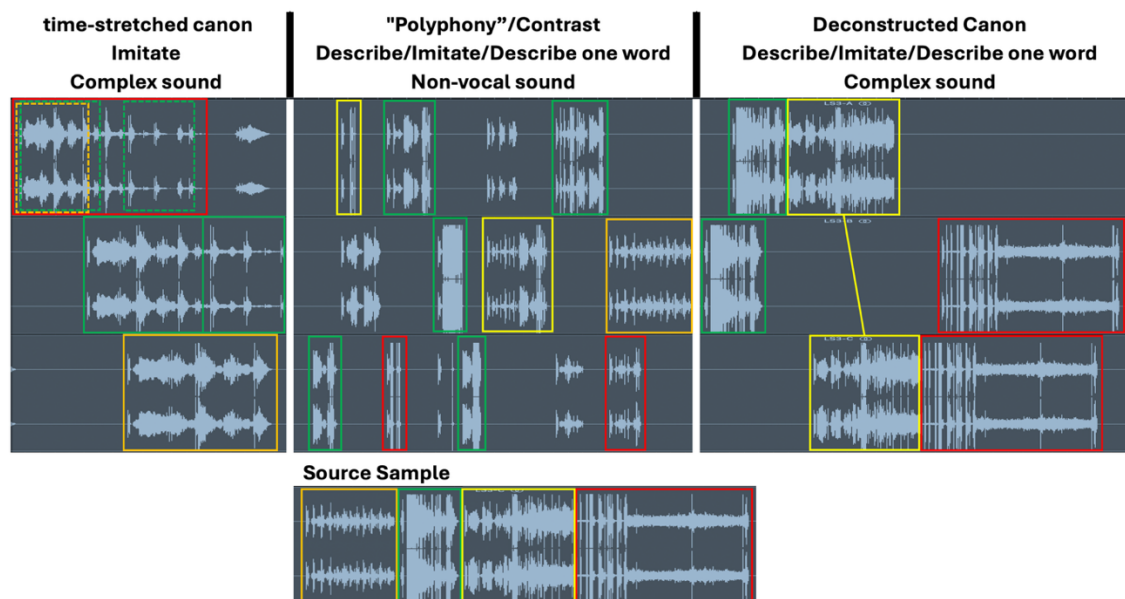


Figure 3. Louis d'Heudieres - *Laughter Studies 3* (2016), Musical Structures in the Central Section: "time-stretched canon; Polyphony Contrast; Deconstructed Canon. The source sample for the materials of Polyphony Contrast and Deconstructed Canon is shown at the bottom of the figure.

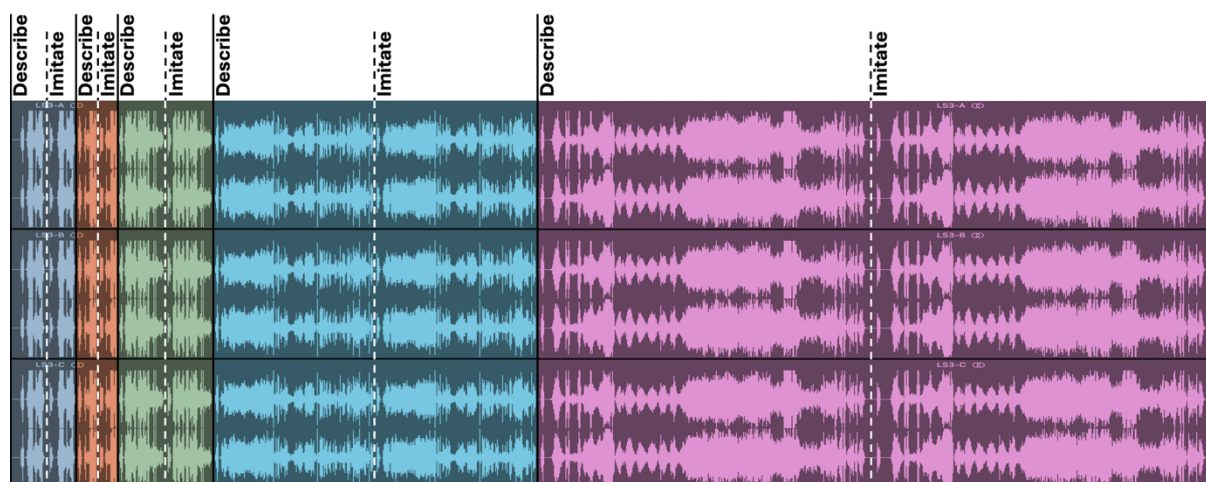


Figure 4. Louis d'Heudière - *Laughter Studies 3* [2016], final “Unison” Section showing the increasing length of the pairs of “describe/imitate” samples.

They are not signifiers in a socially mediated game of linguistic communication, but rather vocal attempts at replicating sonic phenomena” [65]. This abstraction is reinforced by the deliberately non-theatrical staging: performers face the audience, avert their gaze to the tops of audience members’ heads, and maintain poised, confident body language [60], even as they are tasked with vocal actions that progress from challenging to physically untenable (See Figure 3. below).

In the final section, d’Heudière presents a unison or “Tutti”. Each performer hears the same instructions and recorded sounds, but the audience hears something entirely different, the variation in the responses of the three performers to the same “stimuli” – the subject, focus and heart of the work (see Figure 4. above).

Despite its restricted materials - performers receive only three types of instruction and the sonic palette is limited entirely to the human voice - the work achieves considerable structural complexity. Complexity arises in part from d’Heudière’s systematic consideration of every possible “micro-textural” combinatorial relationship between performers and instructions [65], a strategy comparable to John Zorn’s approach in his early Game Pieces. [67]

3.2. Lindsay Vickery: *Un Paso al Abismo* (2018)

Un Paso al Abismo was inspired by the composer’s discovery of a production technique developed by the Mexican media company *Televisa* to accelerate recording schedules for telenovelas (such as *Un Paso al Abismo* (1958-59) (One Step from the Abyss)).

A Mexican engineer invented an electronic earpiece for instant communication with actors that became a standard and somewhat unique element of the *Televisa* production process. Performers could be fed their lines, either between takes or while taping was in process; as a result, the speed of recording was greatly enhanced [68].

The technique was adopted separately in the theatre world in the genre *Verbatim Theatre* that seems have emerged in England and Australia in the early 2000s. Descriptions by Blythe and Oades define an approach that

is remarkably similar to the Audio Score, “a paperless form of theatre that features the faithful reproduction of speech patterns” [69] featuring the use of “spoken instructions, timing cues, or text fragments via headphones during performance”. [70]

Similarly, in *Un Paso al Abismo*, eleven performers imitate, as precisely as possible, excerpts of recorded dialogue from the long-running U.S. soap opera *Days of Our Lives* (1965–), which has aired over 150,000 episodes. [71] The Audio-Score consists of dialogue collages assembled by the composer and selected to foreground what Blake terms the soap’s “latent hysteria”. [72] The material draws on a wide range of *Days of Our Lives* plotlines, principally around the elaborate machinations of the villain Stefano DiMera.

Performers in *Un Paso al Abismo* are required to reproduce the theatrical delivery, timbre, pacing, and dialect of the soap opera actors with high fidelity. Sdraulig and Lortie categorise this use of Audio-Score as “Rehearsed Emphasis / High-fidelity reproduction”, [3] in which performers practise extensively rather than responding to the Audio-Score in real time). This approach enables performers to “execute a high degree of timbral detail in the imitations” [3], but positions them more as a performative conduit rather than an interpreter. The Audio-Score functions less as an improvisatory prompt than as a set of cues or an aide-memoire - closely paralleling the technological mediation of *Televisa*’s telenovela production process.

The Audio-Score parts were assembled directly within a Digital Audio Workstation using samples of dialogue from the soap opera. These samples were selected as archetypal examples of soap opera tropes, such as the following excerpt:

Stefano: “That was too close. (ah, breathes) Abe and Lexi could have ruined my entire plan (breathes) thank goodness nobody knows about this secret tunnel... except for the lady in white.” [61]

The ordering and layering of voices is explicitly collage-like, applying techniques commonly associated with collage practice to musical, verbal, and narrative parameters, including juxtaposition, overlap, interruption, and the fragmentation of sequential dialogue and narrative

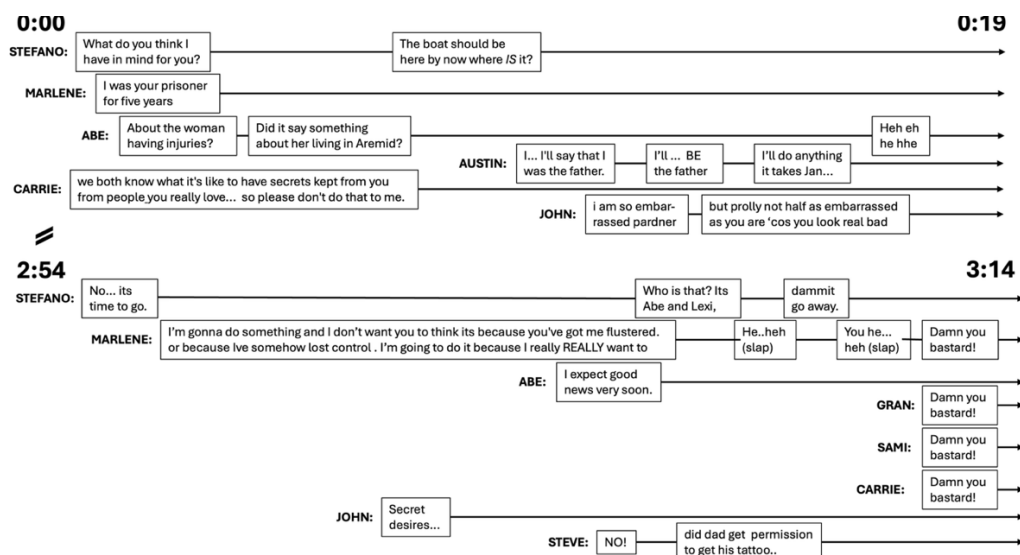


Figure 5. Lindsay Vickery – *Un Paso al Abismo* (2018) (excerpts), transcription of timestamps 0:00-0:19 and 2:54-3:14.

continuity. The opening demonstrates how multiple narrative threads are juxtaposed, relying on the audience’s recognition of dialogue “codes”—for example, those associated with the characters Stefano and Marlena (villain and victim), Carrie (manipulative), and Austin (pathetic pleading)—to infer distinct narrative contexts.

While the work is clearly akin to the *Verbatim Theatre* practices mentioned earlier and has an informal quality reminiscent of “play-back theatre”, like *Laughter Studies* it explores musical/sonic ideas such as the more overtly “musical” orchestration of voices, building toward a crescendo in which four female characters simultaneously cry out, “Damn you, bastard” in the section beginning at 2’54”.

The performance relies on the audience’s capacity to decode these dialogue conventions even within a dense sonic environment of overlapping speech, non sequiturs, and disrupted narrative flow. It also depends on the productive friction between cultural hierarchies: the perceived “low” melodrama of soap opera dialogue is recontextualised within the conventions of the “serious” new music concert, with the camp excess of the performance practice playing a central structural and affective role. Figure 5. presents a transcription of the opening and a later section of the work.

3.3. Elisabeth Schimana - Landscapes (2021) 7x3 Interventions

Landscapes (2021) is part of Schimana’s *Acoustic Memory* series, in which “the musicians are given sound samples as a source material, which they translate to their instruments”. [73] The work

consists of seven roughly three minute “modules”, each created from a simple mosaic of sound objects or “fragments” -either field recordings or electronic samples. It was created for the group *Airborne Extended* that features the unusual combination flute, recorder/ Paetzold, harp and harpsichord. The work was performed and later recorded by *GreyWing* Ensemble in 2022 Although *GreyWing* also has a similarly unusual line-up, the harp part was retained but the other instruments were substituted as follows recorder (flute), flute (clarinet), and harpsichord (electric guitar). The spectrogram examples in Figure 7. below are created from the studio recording.

The 33 “fragments” range from 15 to 75 seconds in duration and must be memorised by each performer. In performance the fragments in each module are recalled/performed in-synch with audio recordings in which they are arranged in a particular configuration of

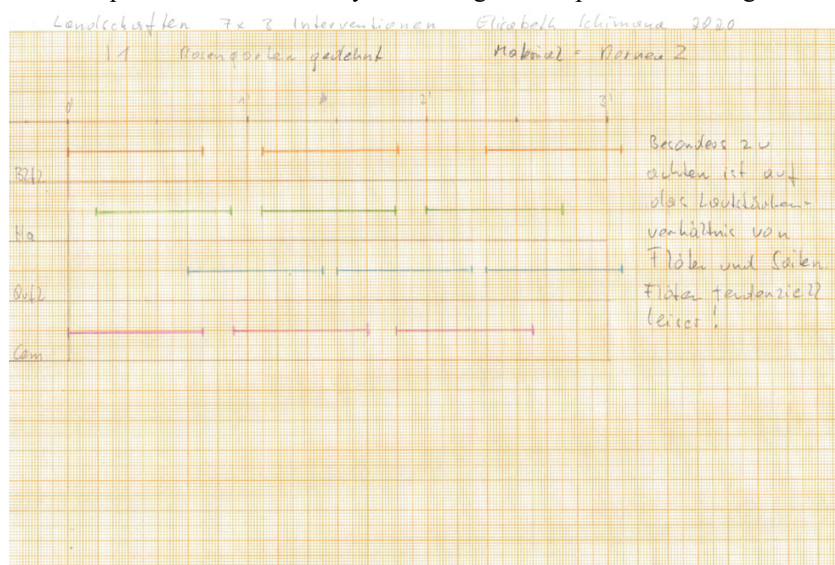


Figure 6. The time structure of Elisabeth Schimana – *Landscapes* (2021) 1L Stretched Rosegarden. The graph grid is in seconds and the durations of the fragments in each instrumental part are represented by different colours recorder/ flute (Orange), harp (Green), flute/clarinet (Blue) and harpsichord/electric guitar (Red).

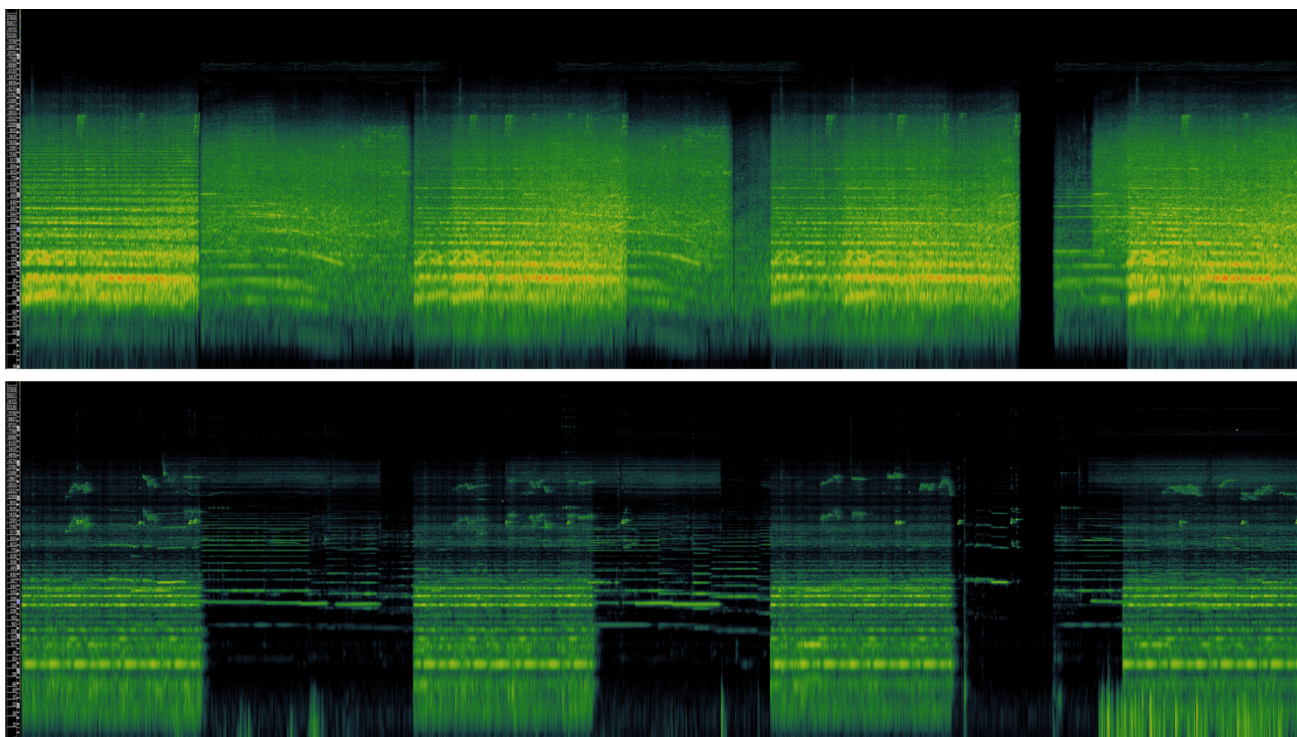


Figure 7. Spectrograms of the electronic audio score (above) and ensemble emulation (below) in *Schimana – Himmel*.

repetitions and silences. The composer provides a “time structure” (See Figure 6.) indicating the temporal relationships between the samples used in each of the four parts.

Landscapes also demands “Rehearsed Emphasis/High-fidelity reproduction”, [3] but the performers here face different challenges from those found in *Laughter Studies* or *Un Paso al Abismo*. Whereas those works all focussed on the challenges and possibilities of the human voice encountering generally different tasks, in *Landscapes* the same or similar tasks are confronted by performers with the idiomatic constraints and possibilities of different instruments.

Rosegarten (Rose garden), the third module of *Landscapes*, is constructed from fragments comprising short electronic sounds, generally spanning a range between ~400Hz to beyond ~20KHz. The majority of the formant energy falls between a spiky 10-20KHz (the composer labels the fragments “dornen” (thorns)). Although such high frequencies may generate artifacts at lower pitches due to aliasing, headphone response and other factors, the highest frequencies are usually inaudible to adults (especially older ones). The actuality of what they physically “sound like” perceptually to each performer is somewhat indeterminate (and slightly existential) and obviously central to a work involving the emulation of audio.

From the performer’s perspective, the capabilities of each instrument play a crucial role. For example, their ranges are quite varied: while the range of the electric guitar is approximately 80 Hz – 5 KHz and the harp 30 Hz -3 KHz, the flute and clarinet are closer to 260 Hz to 2 KHz and 147 Hz to 2 KHz respectively. Predictably the electric guitar, timbrally closest to the recorded sounds, was able to produce the most plausible emulations, the harp slightly

less so, while the flute and clarinet were restricted to a much more impressionistic “low-fidelity” reproduction. However, the variation in performer interpretation is presumably the point in this module, and as is often the case in Audio emulation works, it is the performer’s attempt to emulate the sound that produces the most interesting and unexpected outcomes.

The final module in the set, *Himmel* (Sky), uses a field recording of a helicopter. It is not the generic “sound of a helicopter”, found in sample packs, but rather a real field recording of a slightly distant helicopter - distant “in the landscape”. Emulating the recording, the performers are faced with a different, almost opposite set of problems to those found in *Rose Garden* - the field recording is “noisy”, layered or polyphonic (in the musical sense) and includes low ground hum around 60 Hz, again beyond the range of the flute or clarinet. The harp and the electric guitar with their polyphonic capabilities and wide frequency range cope the most naturally with these challenges. The guitar in particular, able to process its output with a distortion pedal, and perfectly adapted to producing ground hum, was able to create most of the broad frequency noise in the recording (the four periods where the guitar emulates the white noise are clearly visible in contrast to the other instruments’ attempts). The high frequency glissandi on the other hand are more prominent in the live recording than in the recording, partly because of the difficulty for acoustic instruments in creating a “distant” sound live. Figure 7. shows a spectrogram of the audio score (above) and the live performance below).

3.4. Lindsay Vickery: *Little (Grey)Wing* (2022)

Little (Grey) Wing uses samples of materials that are potentially familiar to the audience, from the final two

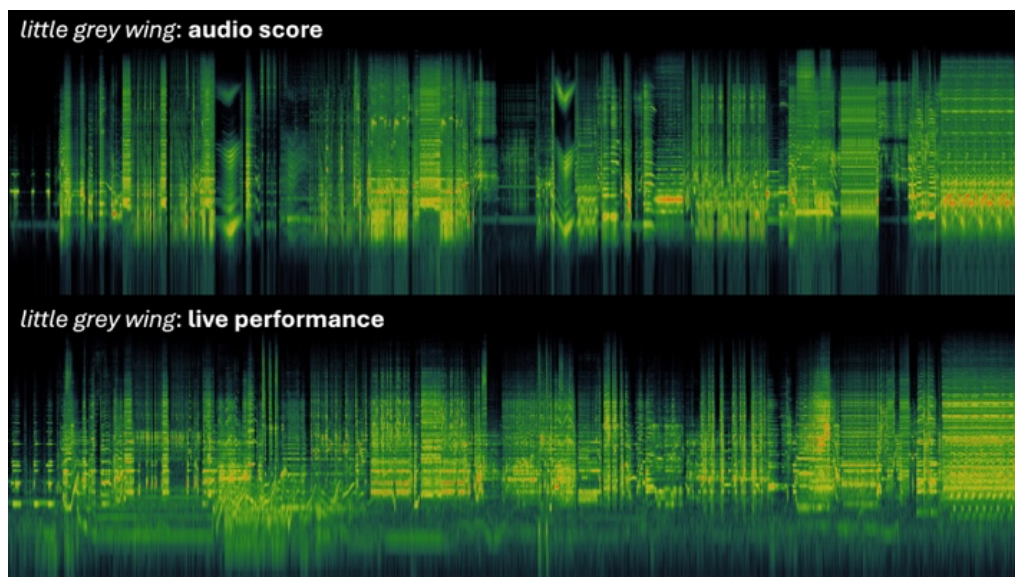


Figure 8, Comparison of Spectrogram of *Little (Grey) Wing* Audio Score (above) and live performance (below).

albums of the *Jimi Hendrix Experience*. The seven-and-a-half-minute Audio Score is a sonic collage of 187 mostly extremely short samples from 21 different tracks (See Table 3.). The samples often consist of fragments of so-called “studio banter”, false starts and doodling that occur between tracks. The title is obviously a nod to both Hendrix – *Little Wing* (1967) and the ensemble performing the work.

The samples are densely distributed, forming a three-part structure in which the first section is stable, the second exploratory and the final section returns again to a more stable but also energetic climax of off-kilter loops. Like *Landscapes*, the work is entirely focused on the instrumental emulation of recorded sound. But unlike the other works in this section, the same audio component is presented to all of the performers in conjunction with a synchronised scrolling score that “orchestrates” the components that they are to emulate. This, along with Rehearsed Emphasis/(slightly tongue-in-cheek) High-

Axis: Bold as Love (1967)

<i>EXP</i>	11
<i>Up From the Skies</i>	5
<i>Ain't No Telling</i>	4
<i>Little Wing</i>	2
<i>If Six Was Nine</i>	37
<i>You Got Me Floatin'</i>	1
<i>She's So Fine</i>	1
<i>One Rainy Wish</i>	1
<i>Bold as Love</i>	10

Electric Ladyland (1968)

<i>Have You Ever Been (To Electric Ladyland)</i>	1
<i>Crosstown Traffic</i>	20
<i>Voodoo Chile</i>	3
<i>Come On (Let the Good Times Roll)</i>	2
<i>Gypsy Eyes</i>	28
<i>Burning of the Midnight Lamp</i>	2
<i>Rainy Day, Dream Away</i>	2
<i>1983 (A Merman I Should Turn to Be)</i>	11
<i>Moon Turn the Tides</i>	2
<i>Still Raining (Still Dreaming)</i>	16
<i>House Burning Down</i>	3
<i>All Along the Watchtower</i>	5
<i>Voodoo Child (Slight Return)</i>	20

Table 3. List of sampled tracks in Lindsay Vickery - *Little (Grey) Wing* (2022) and the number of samples from each.

fidelity reproduction”, [3] potentially contributes to the ensemble’s ability to maintain a close synchronisation, especially rhythmically and timbrally, with the Audio Score collage. This is visible in the comparison between the spectrograph of the Audio Score and the live performance seen in Figure 8.

Although an electric guitar is part of both the original recordings and the live ensemble, the components performed by Hendrix in the recording are often distributed between several performers in the ensemble and highlight the complexity (impossibility) of faithfully emulating the sounds of a rock band with acoustic instruments.

The “orchestration” was strongly influenced by rhythmic and timbral features of the collage rather than the pitched ones. This results in the translation of closed hi hat sounds into key clicks, bass drum strikes into palm strikes on the harps and other extended techniques developed as part of the rehearsal process. This contributes to the performances retaining much of the energy of the original recordings, in a manner that is recognisable but perhaps has something of an “uncanny valley” quality for the audience.

The inclusion of such a detailed score places *Little (Grey) Wing* in the border territory of Audio Score practice, but in performing the work, the primacy of the Audio Score and the visual score’s mnemonic function is clear despite their interdependence.

4. CONCLUSION

This discussion aimed to reframe the Audio score as an approach that reconfigures compositional authority, in particular through its foregrounding of communication through the medium of sound itself and through minimizing the representation of sound via visual abstraction.

The case studies explored, exemplified a diverse range of types of authorial reconfiguration: delegated (*Laughter Studies*), centralised (*Un Paso*), constrained (*Landscapes*) and hybridised (*Little (Grey) Wing*).

It placed the Audio Score at the convergence of new and emerging technological affordances, aesthetic priorities, and listening practices and examined the proto-auditory score works and their contribution to the formulation of the conceptual foundations of the Audio Score.

Boundary works were discussed, demonstrating the necessity for an Audio Score to be defined by reliance on the primacy of auditory communication both in the “outside” case where auditory cueing is employed in the service of a primary score (even if it is indispensable to the performance); and the “inside” case where audio is the primary form of communication even when used in conjunction with other peripheral score materials.

A number of categorisation parameters were explored concerning works in relation to the emergence of supporting technologies. Where possible I have attempted to place these discussions in the framework of my own experience performing Audio Scores – the first of which was surprisingly forty years ago.

Audio Score practice is clearly relatively new, very rich and requires more detailed documentation. Numerous novel combinations of Audio Score capabilities remain to be explored.

5. REFERENCES

1. J. Bell, Audio-Scores: A Resource for Composition and Computer-Aided Performance, Ph.D. dissertation, Guildhall School of Music and Drama, London, U.K., 2016.
2. S. Bhagwati, “Elaborate audio-scores: Concepts, affordances and tools,” in *Proc. Int. Conf. Technologies for Music Notation and Representation (TENOR 2018)*, Montreal, QC, Canada, 2018, pp. 24–32.
3. C. Sdraulig and C. Lortie, “Recent audio-scores: Affordances and limitations,” in *Proc. Int. Conf. Technologies for Music Notation and Representation (TENOR 2019)*, Melbourne, VIC, Australia, 2019, pp. 75–83.
4. E. Schimana, “Sound as score: Live generated audio score and audio score based on acoustic memory,” in *Proc. Int. Conf. Technologies for Music Notation and Representation (TENOR 2019)*, Melbourne, VIC, Australia, 2019.
5. L. P. M. L. d’Heudières, *Scores in Time: Exploring the Use of Sound as a Medium for Notation in Musical Composition*, Ph.D. dissertation, Bath Spa Univ., Bath, U.K., 2020.
6. E. Fuchs *et al.*, “The PAJ casebook: Alcestis,” *Performing Arts Journal*, vol. 10, no. 1, pp. 79–115, 1986.
7. K. Stockhausen, *Kontakte* [score], Universal Edition, Vienna, Austria, 1960.
8. B. Sears, *The Point of Intersection: An Analysis of Interactivity and Synthesis in Historical Models of Mixed Music by Karlheinz Stockhausen and Elaine Lillios*, Ph.D. dissertation, Brandeis Univ., 2023.
9. L. Vickery, “Visualising the sonic environment,” in *Proc. Australasian Computer Music Conf. (ACMC)*, Sydney, NSW, Australia, 2017, pp. 43–50.
10. J. Sterne, *The Audible Past: Cultural Origins of Sound Reproduction*. Durham, NC, USA: Duke Univ. Press, 2003.
11. J. Ashbourn, *Audio Technology, Music, and Media*. Cham, Switzerland: Springer, 2020.
12. D. Morton, *Sound Recording: The Life Story of a Technology*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2006.
13. P. Théberge, *Any Sound You Can Imagine: Making Music/Consuming Technology*. Middletown, CT, USA: Wesleyan Univ. Press, 1997.
14. C. Stankievecch, “From stethoscopes to headphones: An acoustic spatialization of subjectivity,” *Leonardo Music Journal*, vol. 17, pp. 55–59, 2007.
15. C. Neidhöfer, “Bruno Maderna’s serial arrays,” *Music Theory Online*, vol. 13, no. 1, pp. 1–16, 2007.
16. G. S. Kendall, “Juxtaposition and non-motion: Varèse bridges early modernism to electroacoustic music,” *Organised Sound*, vol. 11, no. 2, pp. 159–171, Aug. 2006.
17. L. Cossetini, *Introduction to Luigi Nono: La fabbrica illuminata; Critical Edition*, Milan, Italy: Ricordi, 2010.
18. E. Ghent, “Programmed signals to performers: A new compositional resource,” *Perspectives of New Music*, vol. 6, no. 1, pp. 96–106, 1967.
19. L. Vickery, “Mobile scores and click-tracks: Teaching old dogs,” in *Proc. Australasian Computer Music Conf. (ACMC)*, Brisbane, QLD, Australia, 2010, pp. 63–70.
20. E. Kermit-Canfield, *Spatialization in Selected Works of Iannis Xenakis*, M.A. thesis, Pennsylvania State Univ., University Park, PA, USA, 2013.
21. D. Schiff, *The Music of Elliott Carter*, 2nd ed. Ithaca, NY, USA: Cornell Univ. Press, 1983.
22. S. Haynes, *Musical Thought at IRCAM*, T. Machover, Ed. New York, NY, USA: Taylor & Francis, 1984.
23. A. Schwartz, “Don’t choose the nightingale: Timbre, index, and birdsong in Respighi’s *Pini di Roma*,” in *The Oxford Handbook of Timbre*, Oxford, U.K.: Oxford Univ. Press, 2021.
24. J. E. Higgins, *Nature Music: Prerecorded Nature Sounds in Electroacoustic Composition*, Ph.D. dissertation, Northwestern Univ., Evanston, IL, USA, 2017.
25. M. Kendall, *Sound Works: Model Listeners in Soviet Art, 1929–1941*, Ph.D. dissertation, Univ. of California, Berkeley, CA, USA, 2019.
26. L. Victoria, “Avant-garde machines: On the integration of technology and art,” *Technology and Language*, vol. 6, no. 3, pp. 80–96, 2025.
27. M. Katz, “Hindemith, Toch, and *Grammophonmusik*,” *Journal of Musicological Research*, vol. 20, no. 2, pp. 161–180, 2001.
28. J. Pritchett, *The Music of John Cage*. Cambridge, U.K.: Cambridge Univ. Press, 1996.
29. J. M. Green, “*Available to Our Ears*”: *John Cage and Broadcast Media*, Ph.D. dissertation, Univ. of Rochester, Rochester, NY, USA, 2020.
30. P. Schaeffer, *In Search of a Concrete Music*. Berkeley, CA, USA: Univ. of California Press, 2012.

31. L. E. Hartman, *DIY in Early Live Electroacoustic Music: John Cage, Gordon Mumma, David Tudor, and the Migration of Live Electronics from the Studio to Performance*, Ph.D. dissertation, Louisiana State Univ., Baton Rouge, LA, USA, 2019.
32. C. Salter, *Entangled: Technology and the Transformation of Performance*. Cambridge, MA, USA: MIT Press, 2010.
33. T. Patteson, *Instruments for New Music: Sound, Technology, and Modernism*. Berkeley, CA, USA: Univ. of California Press, 2015.
34. C. Kelly, *Cracked Media: The Sound of Malfunction*. Cambridge, MA, USA: MIT Press, 2009.
35. K. Stockhausen, *Spiral für einen Solisten (Spiral for a Soloist)*, Nr. 27 [score]. Vienna, Austria: Universal Edition, 1973.
36. R. Erickson, *Pacific Sirens* [music score]. Smith Publications, 1968.
37. J. Mackay, "On Community, Institutions and Politics in the Life and Work of Robert Erickson". in *Music, Metamorphosis and Capitalism: Self, Poetics and Politics*. Cambridge Scholars Publishing p. 27, 2007.
38. R. Erickson, *Sound Structure in Music*. Berkeley, CA, USA: Univ. of California Press, 1975.
39. G. Bryars, *I, 2, 1–2–3–4* (1971) [score], in *Verbal Anthology*. London, U.K.: Experimental Music Catalogue, 1972.
40. G. Bryars, "1, 2, 1–2–3–4," liner notes, in *Ensemble Pieces*. London, U.K.: Obscure Records, 1975.
41. J. Talman, (2000). Joshua Fried: Subverting Technology <http://www.mvdaily.com/articles/2000/02/emfried4.htm> [online]
42. M. Á. Melgares y A. Hampton, "Entrevista con Ant Hampton," *SOBRE: Prácticas artísticas y políticas de la edición*, no. 7, pp. 122–130, 2021.
43. A. Lucier, *Carbon Copies* [score]. New York, NY, USA: Material Press, 1989.
44. A. Lucier, *Memory Space* (1970), in *Chambers*, A. Lucier and D. Simon, Eds. Middletown, CT, USA: Wesleyan Univ. Press, 1980.
45. L. Marchetti, "The coast and the coast opposite," liner notes, Bandcamp, 2018. [Online].
46. S. Hosokawa, "The Walkman effect," *Popular Music*, vol. 4, pp. 165–180, 1984.
47. J. Wilson, "40 years of the four-track," *Engineering & Technology*, vol. 14, no. 9, pp. 46–47, 2019.
48. C. Yavelow, "MIDI and the Apple Macintosh," *Computer Music Journal*, vol. 10, no. 3, pp. 11–47, 1986.
49. E. Jones, "Macintosh or Atari?" *Sound On Sound*, Jan. 1988, pp. 25–30.
50. A. Ballista, E. Casali, J. Chareyron, and G. Haus, "A MIDI/DSP sound processing environment for a computer music workstation," *Computer Music Journal*, vol. 16, no. 3, pp. 57–72, 1992.
51. K. S. Ladányi, "Digital sequencers: A compositional tool for general music," *General Music Today*, vol. 8, no. 3, pp. 20–24, 1995.
52. C. Roads, *The Computer Music Tutorial*. Cambridge, MA, USA: MIT Press, 1996.
53. A. Camurri and P. Ferrentino, "Interactive environments for music and multimedia," *Multimedia Systems*, vol. 7, no. 1, pp. 32–47, 1999.
54. J. K. Downs, *Headphone Listening: Space, Embodiment, Materiality*, Ph.D. dissertation, Univ. of Sheffield, Sheffield, U.K., 2021.
55. P. Roquet, "Acoustics of the one-person space: Headphone listening, detachable ambience, and the binaural prehistory of VR," *Sound Studies*, vol. 7, no. 1, pp. 42–63, 2021.
56. J. J. Arango, P. A. Barbosa, and E. V. Astorga, "Headphone listening cultures: Contingencies and crossings," in *Proc. Sonologia 2020*, São Paulo, Brazil, 2020, pp. 473–492.
57. R. Capretta, *The Complexity of Sound Design and Operations for Television Production*, M.A. thesis, McGill Univ., Montreal, QC, Canada, 1994.
58. G. Jones, D. Layer, and T. Osenkowsky, "Intercom and IFB systems," in *NAB Engineering Handbook*, 10th ed., E. Williams, Ed. New York, NY, USA: Routledge, 2007, pp. 485–486.
59. C. Hope, *Tough It Out* [score instructions]. Perth, WA, Australia, 2014.
60. L. P. M. L. d'Heudières, *Laughter Studies 3* [score]. Self-published, 2016.
61. L. Vickery, *Un Paso al Abismo* [score]. Self-published, 2018.
62. E. Schimana, *Landscapes* [musical work]. Self-published, 2021.
63. L. Vickery, *Little (Grey) Wing* [score]. Self-published, 2021.
64. J. Saunders, "Perfect information: Scores as self-evident processes," in *Proc. Int. Conf. Technologies for Music Notation and Representation (TENOR 2024)*, Zurich, Switzerland, 2024, pp. 53–57.
65. L. P. M. L. d'Heudières, *Scores in Time: Exploring the Use of Sound as a Medium for Notation in Musical Composition*, Ph.D. dissertation, Bath Spa Univ., Bath, U.K., 2020.
66. C. Sdraulig and L. P. M. L. d'Heudières, "Attending to attending: Compositional attention and mediated listening," *Tempo*, vol. 76, no. 300, pp. 18–32, 2022.
67. H. Mandel, *Future Jazz*. Oxford, U.K.: Oxford Univ. Press, 1999.
68. A. Paxman, "Hybridized, glocalized and *hecho en México*," *Global Media Journal*, vol. 2, no. 2, 2003. [Online].
69. R. Oades, "Creating a headphone-verbatim performance," in *Verbatim: Staging Memory and Community*. London, U.K.: Oberon Books, 2010, pp. 84–87.
70. D. Prudhon, "Alecky Blythe and 'headphone verbatim,'" in *The New Wave of British Women Playwrights*, vol. 33. London, U.K.: Palgrave Macmillan, 2023, p. 205.
71. E. Lewis, "Days of Our Lives renewed for season 61 at Peacock," *Soap Opera Network*, Nov. 1, 2024. [Online].
72. L. Blake, "Trapped in the hysterical sublime," in *Return to Twin Peaks*. London, U.K.: Palgrave Macmillan, 2016, pp. 229–245.
73. E. Schimana, *Landscapes*, <https://elise.at/projekt/Landschaften>, [online], 2021.

LISTENING GUIDES AS COMPOSITIONAL DEVICES: TEXTUAL AND FORMAL STRATEGIES IN *LES DISPARITIONS*

Sébastien Roux

Ircam

sroux@ircam.fr

ABSTRACT

This article examines the notion of the listening guide, considering it not as an external device accompanying a work, but as an operator integrated into the very form of the sonic experience. It is based on *Les Disparitions (ambiantes et acousmatiques)*, a project taking the form of a listening session with spatialized sound, organized around the idea of sound disappearance within the context of field recording. The project articulates two complementary modes of listening guidance. The first relies on a text-based guide synchronized with the music and delivered to the audience via their mobile phones, momentarily orienting auditory attention. The second is embedded within the sonic form itself, where the progressive exposure and subsequent disappearance of a *target sound* bring about a transformation in listening modes. Based on a research protocol grounded in listening interviews, this article shows how these strategies of disappearance foster a transition from causal listening toward a mode of listening oriented toward the sonic properties themselves.

1. INTRODUCTION

A listening guide generally refers to a set of materials intended to support the understanding of a work. It can take various forms—texts, program notes, interviews, contextual information—and brings together heterogeneous elements made available to the listener in order to illuminate their experience. Musicologist Nicolas Donin emphasizes the dual function of the listening guide, which constitutes both “information about the work and a training of the ear,” adding that “the term information must therefore be understood in a strong sense: listening guides give form to the attentive listening they prescribe” [1]. Through familiarity with these elements, listening can be transformed: attention shifts, concentrates, or is oriented toward specific dimensions of the work, thereby reinforcing the appreciation of its aesthetic or formal qualities.

In most cases, however, the listening guide is consulted outside the work itself. It prepares, shapes, and extends the concert experience through a consultation that usually takes place beforehand. By considering the listening guide as an orienting device for attention, this article presents

an artistic form that integrates the listening guide into the very conception of the work, making it synchronous with the sonic unfolding. In other words, the listening guide acts in real time on the listener’s attention and becomes an integral part of the listening situation.

This proposal is explored through two complementary modes of guidance. The first, explicit, takes the form of a text-based guide distributed to the audience via their mobile phones and synchronized with the sonic unfolding, acting directly on auditory attention. The second, implicit, relies on a formal organization of sound that progressively orients attention and functions as a listening guide.

These two dimensions converge in a hybrid form, distributed between text and sound: *Les Disparitions (ambiantes et acousmatiques)*¹. The project takes the form of an hour-long listening session of multichannel field recordings diffused through an acousmonium. The piece is presented as a collection of nine movements staging the disappearance of a sonic environment or of a *target sound* within an environment. Initially exposed, the sound is gradually eroded, masked, erased, or even forgotten. The piece was commissioned by the Archipel Festival (Geneva) and premiered in April 2025.

In this article, I present the modalities of these two forms of guidance, the levels at which they operate, and the ways in which they transform the listener’s attention.

2. PROJECT DESCRIPTION

Les Disparitions is a project grounded in the practice of field recording, a practice that “is now commonly encountered both as a distinct artistic practice and as a component of experimental music and sound art” [2]. *Les Disparitions* is structured around a central notion: disappearance. This notion should be understood as the erasure or masking of a sound source—the starting point of the disappearance process—within a sonic environment, or as the progressive alteration of the sonic environment itself. Whether it is a punctual sound source or a more global sonic environment, the sound destined to disappear is referred to here as the *target sound*.

Different strategies are employed to implement these disappearances. The dominant strategy is distancing: the *target sound* is recorded from the closest to the most distant position, according to a process of either continuous or discrete (organized in steps) distancing. This distancing results in a progressive loss of energy and precision in the

Copyright: ©2026 Sébastien Roux et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ The Disappearances (ambient and acousmatic).

No.	Target sound	Strategy	Recording technique
1	Canal lock	Discrete distancing (7 steps)	Hexaphonic (star configuration)
2	Waves on a pebble beach	Discrete distancing (6 steps)	Hexaphonic (linear array)
3	Highway	Discrete distancing (6 steps)	Hexaphonic (star configuration)
4	Vent opening in an urban street	Attentional diversion through the progressive emergence of the surrounding sonic environment	Ambisonic + microphone on the vent opening
5	Church bells	Attentional diversion	Hexaphonic (star configuration) + cardioid microphone
6	Cello playing a repeated pattern	Continuous distancing	Hexaphonic (star configuration)
7	Clarinet playing a sustained tone	Continuous distancing and masking	Hexaphonic (star configuration)
8	Crowd noise from a stadium during a sports event	Discrete distancing (6 steps) and masking	Ambisonic
9	Wind noise in a pine forest	Attentional diversion through verbal description of sounds absent from the recording	Hexaphonic (star configuration)

Table 1. List of the nine movements in *Les Disparitions*.

target sound, leading to its perceptual disappearance.

A second strategy is based on masking: the *target sound* becomes covered or buried by another environmental sound. This process may be combined with distancing, as the *target sound* progressively recedes and eventually becomes masked by a new sound.

A third strategy involves attentional diversion. In this case, the *target sound* does not necessarily disappear from the sonic environment itself, but the listener's awareness of its presence—and the orientation of attention toward it—gradually fades as other sonic or textual events capture attention. The listener only becomes aware of the disappearance of the *target sound* after it has already occurred.

The disappearances are produced without the use of mixing or equalization. The perceptual transformations rely exclusively on choices of sound recording, editing, and temporal organization of the recorded materials. Each disappearance is produced using a multi-microphone recording setup enabling spatialized reproduction. Two main techniques are employed: ambisonic recording and hexaphonic recording using six microphones arranged in a star or line configuration.

For each recording session, sound levels are measured and subsequently reproduced during diffusion, so that playback levels remain as close as possible to the original recording situation.

The project currently consists of nine movements, summarized in Table 1, which indicates for each movement, a *target sound*, a disappearance strategy, and a specific recording technique.

The nine movements have been presented in the form of a collective listening session with diffusion over an acousmonium. The acousmonium used for the performance consisted of a heterogeneous and spatially distributed loudspeaker system, combining studio monitors, sound reinforcement speakers, and unconventional sources (horns, piezo elements). Loudspeakers were deployed across multiple spatial layers—including the stage, around and be-

hind the audience, as well as at elevated positions (balcony, catwalk, and ceiling)—forming an immersive and multidirectional diffusion environment. The piece was presented in multichannel playback without live interpretation (in the sense of the GRM tradition), with spatial balances and levels pre-calibrated during rehearsals.

3. TEXTUAL GUIDANCE

Textual guidance is based on a communication system in which a server sends text messages to the audience's mobile phones, ensuring synchronized display during the listening session.

Before the beginning of the piece, audience members are invited to connect to a web page via their mobile phones. They are informed that, during the session, information about the sound recordings will be displayed on their screens. In order to prevent spectators from keeping their eyes fixed on their phones while waiting for the next message, a light signal indicates when text appears. This signal takes the form of a slow and gradual transformation of light—perceptible yet non-intrusive—which functions as an attentional operator designed to preserve continuity of listening. By indicating precisely when to look at the screen, it relieves the listener from the need to monitor their device continuously, allowing them to remain primarily engaged in listening.

Other communication channels were initially considered, such as a pre-recorded voice superimposed on the sound recordings (as in *Presque Rien No. 2 ainsi continue la nuit dans ma tête multiple* by Luc Ferrari), or a projection on a screen placed at the center of the stage. These options were ultimately rejected. In the first case, the superposition of an additional sonic element would introduce an excessive disturbance to the listening process. In the second case, the presence of a single screen toward which all gazes converge risks turning the listening session into a film screening, producing a form of externalization of listening.

The use of the mobile phone instead allows for an individ-

	1	2	3	4	5	6	7	8	9
Title	B	B	B			B	B	B	
Target sound	B	B	B	D	D	B	B	B	D
Localization	B	B	B	D	D	B	B	B	D
Dating	B	B	B		D	B	B	B	
Strategy	B	B	B			B	B	B	
Extra contextualization				D	D				
Interrogative register			B				A		
Explicit listening instructions							B		
Description of the phenomena			B	D	D		A		D
Explicit presence of the narrator			B						D

Table 2. List of the different guidance modes in the nine movements in *Les Disparitions*. Text is read before listening (B), during listening (D), after listening (A) to the corresponding movement.

ualized address, fostering a more intimate mode of listening. In this respect, it is reminiscent of *An Hour for Piano* by Tom Johnson, a solo piano work accompanied by a written text silently read by the audience during performance, describing in the first person the performer’s actions and the temporal structure of the piece, with the stated aim of “increasing your ability to concentrate on the piece, and not to distract from it”².

The textual content displayed on the audience’s phones consists of sentences, short text sequences, or isolated words, whose appearance is precisely controlled in time. Display parameters—duration, character appearance rate, fades—are integrated into the composition and allow for a gradual dissemination of the text. Some texts appear just before the listening of a movement (approximately 70% of the text), functioning as a title; others appear during listening (approximately 25%), others appear after the listening of a movement (approximately 5%). Text segments that were not presented during listening (i.e., before or after) not only conveyed information but also contributed to structuring the overall form, creating an alternation between focused listening and moments of pause or release between movements. The text segments presented during listening introduced a multimodal experience combining reading and listening.

The textual guidance encompasses several types of utterances. Some messages function as titles (“Toward background noise,” “Exogenous sounds”) and offer a general description of the disappearance process. They may be associated with localization and dating information (“Haute-Savoie, 2024”, “Recorded on Boulevard Carl Vogt in Geneva”). At times, explicit information about the disappearance strategy is provided (“Strategy: 6 steps,” “the microphone moves away on a conveyor belt”). Other messages provide contextualization by precisely situating the recording (“At the corner of Patru Street, there is a vent opening at the sidewalk level. Standing above it, one can distinctly hear the sound of a machine”). The text may also adopt an interrogative register, sometimes very direct (“Was this the case for you?” or “If you had to rate the degree of similarity between the final stage of the waves and that of the

lock, what would your answer be?”), or propose explicit listening instructions (“If you keep your attention focused on the sound of the clarinet, can you identify the moment when you stop hearing it?”). One also finds descriptive formulations of the sonic phenomenon as it unfolds (“It is a tenuous sound, it blends into the landscape and then disappears,” “Both modulate, stop, then resume”), as well as the explicit presence of a first-person “I”, establishing a relationship of trust and familiarity with the listener (“I suppose you would say that the similarity is rather high,” “The recording I made that day is longer”). The guidance modes are summarized in Table 2.

The movements based on distancing strategies (1, 2, 3, 6, 7, 8) rely on textual information delivered prior to listening, whereas movements based on attentional diversion (4, 5, 9) use textual guidance during listening. This difference can be explained by the degree of perceptual clarity of the process at work. In the case of distancing, the process is highly explicit: the *target sound* is recorded at increasing distances until it eventually disappears. In contrast, attentional diversion requires more subtle intervention; in order to successfully redirect attention away from the *target sound*, auditory attention is guided through suggestions delivered during listening.

The text delivered via mobile phone presents the recording context and outlines the process employed for each disappearance. This information allows the listener to form a clearer understanding of both the form of each movement and the overall structure of the piece. By addressing these formal questions in advance, the listener can devote fuller attention to the sonic properties of the disappearances. Questions, listening instructions, and descriptive formulations orient the listener’s attention toward specific sonic qualities, propose attentional postures that might not be adopted spontaneously, and foster a reflexive stance toward one’s own listening. In this sense, the textual guide can be understood as fostering what Barry Truax calls listening-in-readiness: an intermediate mode of listening in which attention is not fully focused on sound, yet remains available to receive and evaluate potentially significant information. Rather than directing attention toward the identification of sources, the guide helps maintain a state of attentive openness to the evolving sonic environment [3].

² Program notes by Tom Johnson.

4. DISAPPEARANCE STRATEGIES AND CONSEQUENCES FOR AUDITORY ATTENTION

In this section, I suggest that the very form of the movements composing *Les Disparitions* functions as a listening guide. In what follows, we focus more specifically on the movements that employ strategies of discrete distancing. By first exposing the target sound in the clarity of its source, the piece engages what is perhaps the most spontaneous listening attitude: causal listening, oriented toward the search for cues pointing to the originating event. By then progressively distancing it until the source becomes unrecognizable, it creates the conditions to loosen the perceptual link described by Michel Chion—the link that instinctively ties the ear to the material origin of sound—and to invite the listener to focus on the morphological properties of the sound object itself, independently of its provenance [4]. This process tends toward a transition from one mode of listening to another, echoing Schaeffer's distinction between causal and reduced listening [5]. It should be noted that for each movement employing distancing, the nature of the source is given in the title. Thus, rather than avoiding causal listening, the piece deliberately establishes it as a starting point, in order to observe its possible transformation over time.

To test this hypothesis, we conducted an experiment based on a series of listening interviews, focusing on the third movement of *Les Disparitions*. The experimental protocol presented below does not aim to evaluate the work as such, but rather to investigate one specific compositional intuition emerging from it: the hypothesis that progressive distancing may reorient listening from causal identification toward attention to sonic properties.

4.1 Presentation of the third movement of *Les Disparitions*

During the listening session, this movement is presented to the audience using the following text:

In the Background Noise 3
Target sound: Highway
Strategy: 6 steps
Haute-Savoie, 2024

This disappearance takes the form of six successive sound recordings (labeled *highway1* to *highway6* in the experiment), separated by five seconds of silence. All recordings were made on the same day, October 29, 2024. The first recording (*highway1*) was made on a bridge over the highway, approximately five meters above the roadway. The measured sound level at this point was 82 dBA.

Subsequent recordings were made while progressively moving away from the highway, following a protocol in which the average sound level of each step decreases by approximately 6 to 10 dBA relative to the previous one, reaching 42 dBA for the final recording (*highway6*).

The first recording presents clearly individuated sonic entities (vehicles), characterized by rapid dynamic envelopes (marked crescendos and decrescendos) and associated with fast spatial movements. The final recording, by contrast,

presents a noisy sonic texture with no perceptible individuality, minimal spectral or dynamic variation, and no marked spatial movement. The intermediate recordings form a gradual transition between these two perceptual regimes, shifting from a listening mode based on individuated sources to a perception of sound as a global material. It is important to note that although the six recordings all document the same highway, their sonic qualities differ markedly in terms of level, dynamics, spectral content, and spatial movement.

4.2 Methodology

Twelve participants (listed as P1, P2, P3 ... P12) took part in individual listening sessions. The profiles of the participants are listed in Table 3. This relatively small sample should be understood within the framework of an exploratory, qualitative approach. The aim was not to establish generalizable results about listening behavior, but rather to document and analyze tendencies observed within a specific listening context. Moreover, given the participants' familiarity with sound practices, their responses may reflect forms of listening shaped by prior experience. The observations presented here should therefore be considered as situated and indicative, rather than representative of listening in general.

The entire experiment was conducted through headphones using a binaural rendering of the recordings. At the beginning of each experiment, the listener was informed that they would hear a highway recording. The session began with listening to the final stage (*highway6*). This listening was followed by an « explication interview » focusing on the listening experience [6].

Participants were then informed that they would hear a sequence composed of six recordings of a highway, from the closest to the most distant. Although this prior information may orient participants toward a specific referent, it was intentionally provided in order to reproduce the conditions of the concert situation, in which the source is explicitly identified. They then listened to the entire sequence (*highway1* to *highway6*) without interruption, exactly as in the concert situation, with the exception that listening took place through headphones rather than through an acousmionium.

Finally, an explication interview was conducted sequence by sequence: for each recording, a short excerpt was replayed, and the participant was asked to describe precisely what they had perceived during that listening.

The primary objective of this experiment was to determine whether the listening experience of the final stage (*highway6*) within the full sequence differs from that of the same recording heard in isolation, in other words, whether listening to the disappearance process informs and transforms the final listening experience.

4.3 Analysis of listening interviews

When listening to *highway6* heard in isolation (i.e., without prior exposure to *highway1* through *highway5*), participants primarily described an experience oriented toward identifying the sound source. For instance, P1 attempted

Participant	Profile	Use field recordings in their own practice	Age	Gender
P1	Ethnomusicology student	Yes	47	F
P2	Composer	Yes	49	M
P3	Bachelor of Arts student	No	26	F
P4	Sound designer	Yes	30	M
P5	Improviser	Yes	57	M
P6	Electroacoustic music student	Yes	25	M
P7	Bachelor of Arts student	No	22	F
P8	Bachelor of Arts student	No	22	F
P9	Bachelor of Arts student	No	20	M
P10	Electroacoustic music student	Yes	34	M
P11	Music composition student	No	30	M
P12	Electroacoustic music student	No	29	M

Table 3. Participant profiles and relationship to field recording practices.

to situate the recording by imagining their own position, the time of day, the quality of the light, or the location of the person making the recording.

Although the highway is very distant and often perceived as a diffuse mass or background noise, listening remains largely causal. Listeners seek to verify the nature of the sound, often doubting its origin. Some wonder whether the sound might be fabricated or artificially produced (“I wondered whether it might be a reproduction, or a synthetic recreation of a sonic environment” (P12)), while others hesitate between different possible sources, such as the sea or a waterfall (“you wonder where things come from—whether it’s nature, whether it’s cars,” (P9)).

Listening to *highway1* is described as a clear immersion into the highway itself: “here, we are really inside it” (P12). Each vehicle is perceived as an individual entity (“you can hear each car producing a different sound” (P12)), with particular attention paid to movement and spatial dynamics: “what caught my attention was the movement, the feeling that things are arriving and then leaving” (P6). This listening is sometimes experienced as overwhelming or stressful with participants describing a sense of physical danger or discomfort: “It was stressful. There is something a bit dangerous that physically stresses me, I want to step back” (P11).

The sequences *highway2*, *highway3* and *highway4* mark a gradual transition from the perception of individuated sonic entities toward a more global apprehension of the source. Participants describe how “the cars started to blend together” (P11), how “the entities begin to merge, without disappearing completely” (P12), or how “one starts to discern individual sounds less clearly; everything is more blended into a kind of constant murmur” (P7). During this phase, the sound increasingly takes the form of a continuous mass rather than a succession of distinct events.

Finally, when listening to *highway5* and *highway6*, the sound is primarily perceived through its sonic properties rather than its causal origin, and may even be interpreted in terms of musical genre (*highway5*, for instance, being described as a “drone” by P2). Although the sonic material of *highway6* heard within the full sequence (*highway1*

through *highway6*) is identical to that of *highway6* heard in isolation, the perceptual experience is markedly transformed. Interviews show that listening largely shifts away from questions of origin toward a focus on the sonic material itself. Some participants explicitly refer to “sonic matter” and note the disappearance of cars as perceptual objects (P5). Several listeners describe the sound as an ambiguous, almost autonomous sonic object, detached from its referential source: “I’m more in my ears than in the projection of an imagined place” (P7).

This listening trajectory—by progressively stabilizing the nature of the source—enables listeners to devote greater attention to the sound’s intrinsic properties. For some participants, this heightened state of attention, induced by the protocol itself, becomes explicitly conscious. Comparing their listening experiences of *highway6* heard in isolation to *highway6* heard within the full sequence, P12 noted that they “became more absorbed, focusing on the sound itself rather than anything else—really on the sonic material” in the latter case.

Several explanations could account for this shift, including habituation, prolonged exposure, or the progressive abstraction of the sonic material. However, explanations based solely on habituation or prolonged exposure can be questioned, since—even though the source remains the same (a highway)—the acoustic properties of the recordings vary significantly from one sequence to another. While the present protocol does not allow these hypotheses to be fully disentangled, several elements suggest that the formal trajectory itself plays a structuring role.

This analysis suggests a tendency toward a reorientation of listening, moving away from a strictly causal mode toward a more informed listening posture. Rather than a unidirectional shift, the listening process appears to involve a dynamic oscillation between modes, with listeners occasionally returning to source identification, albeit less persistently than in the initial stages. Over time, attention becomes increasingly oriented toward the sonic properties of the sound itself.

5. CONCLUSION

This article proposes considering the listening guide as an operator integrated into the work itself. Taking *Les Disparitions (ambiantes et acousmatiques)* as an example, it shows how the combination of synchronous textual guidance delivered via mobile phones and an implicit form of guidance embedded in the sonic form contributes to shaping auditory attention and, consequently, transforming the listening experience. In other words, and more generally, we observed that “(...) the information listeners receive before hearing a piece of music can significantly alter their experience; this information can not only bias their evaluation but, more interestingly, prompt them to adopt specific listening behaviors, ultimately leading to different aesthetic qualities being ascribed to the music heard” [7].

By integrating contextual elements and proposing specific listening modalities, the project outlines an approach to field recording that moves away from both naturalism and contemplative listening. Instead, it embraces the compositional dimension of the practice and invites a reflexive attitude toward listening. Far from revealing unheard sounds, *Les Disparitions* encourages a reconsideration of how we listen to everyday sounds and our sonic environment.

Acknowledgments

Marie Jeanson commissioned the piece for the Archipel Festival (Geneva 2025). The mobile interface was programmed by Benjamin Matuszewski (Ircam). Hexaphonic microphone was designed by Loïc Lachaize. Recordings

were made with the help of Rudy Décélière, Olivier Renouf, DD Dorvillier. Maxime Guitton, Pierre-Laurent Cassière, Jean-Luc Gergogne and Olivier Meyer coordinated the experiments. I am very grateful to the participants who took part in the experiments. I would like to thank Clément Canonne (Ircam) for his trust and careful advice.

6. REFERENCES

- [1] N. Donin, “Pour une “écoute informée” de la musique contemporaine : quelques travaux récents,” *Circuit. Musiques contemporaines*, vol. 16, no. 3, pp. 51–64, 2006.
- [2] W. Montgomery, “Beyond the soundscape: Art and nature in contemporary phonography,” in *The Ashgate Research Companion to Experimental Music*, J. Saunders, Ed. Ashgate Publishing, 2009, pp. 145–162.
- [3] B. Truax, *Acoustic Communication*, 2nd ed. Ablex Publishing, 2001.
- [4] M. Chion, *Le son*, 2nd ed. Armand Colin, 2010.
- [5] P. Schaeffer, *Traité des objets musicaux*. Seuil, 1966.
- [6] N. Marty, *Quand l’écoute fait l’oeuvre : L’étude des conduites d’écoute de 1979 à aujourd’hui*. Delatour, 2020.
- [7] C. Canonne, “Listening to improvisation,” *Empirical Musicology Review*, vol. 13, no. 1-2, pp. 2–15, 2019.

LIVING AUDIO SCORES: DESIGNING LISTENING-BASED NOTATION IN NADA SFERA PLAYGROUND

Ainolnaim Azizol

Faculty of Music
Universiti Teknologi MARA Malaysia
ainolnaim7683@uitm.edu.my

ABSTRACT

Nada Sfera playground is a web-based live soundscape installation in which environmental audio streams are transformed into musical form through active listening and real-time interaction. Rather than positioning sound as a fixed material to be reproduced or annotated, the system invites visitors to engage with sound as a dynamic, evolving process. Participants shape the sonic outcome by manipulating parameters such as real-time granular synthesis, spatial direction, sound diffusion, and acoustic intensity, allowing each interaction to generate a unique realisation of the work. As the sonic environment is continuously variable and non-linear, performers and visitors require a form of guidance that supports interpretation without prescribing exact outcomes. In response, this paper proposes the concept of living audio scores as a listening-based notational framework for Nada Sfera Playground. These scores function as perceptual roadmaps rather than visual or symbolic representations, guiding musical form through attention, spatial awareness, and textural listening. Where visual or symbolic elements are present, they operate only as secondary orientation cues and do not function as a visual score. By articulating macro-structural trajectories while allowing micro-level sonic variation, living audio scores enable coherent yet open-ended performances within a live soundscape context. The study situates Nada Sfera Playground within contemporary discourse on sound as score, non-visual-focused notation systems, and ecological listening, arguing that listening-based notation offers a viable alternative to reading-oriented score paradigms in web-based and installation-based sonic practices.

1. INTRODUCTION

Musical notation has historically functioned as a visual system designed to encode sound for later interpretation

through reading. From staff notation to ‘graphic’ and animated scores, the act of performance has typically been mediated by visual presentations that prescribe pitch, rhythm, structure, or gesture. While such systems have been extended and challenged throughout the twentieth and twenty-first centuries, particularly within experimental, electroacoustic, and sound art practices, notation remains largely oriented toward visual decoding. In contexts where sound itself is fluid, contingent, or environmentally situated, this reliance on visual representation can become limiting, privileging fixed outcomes over perceptual engagement and reducing listening to a secondary role.

Recent developments in sound-based practices have increasingly foregrounded listening as an active, interpretative act rather than a passive reception of pre-defined material [1, 2]. Within soundscape composition, ecological sound art, and live electronic performance, sound often unfolds as a process shaped by environmental conditions, performer interaction, and system behaviour [3, 4]. In such contexts, the question is no longer how sound should be represented visually, but how musical form can be guided, navigated, and articulated through listening itself.

Within this broader discourse, the notation of sound as score has emerged as a critical framework for understanding how sonic material can function directly as a structuring and instructive element, particularly in live electronic practices where form is realised through interaction and listening [5]. Rather than serving as an object to be notated, sound becomes the primary carrier of musical information, shaping form through its temporal, spatial, and textural properties. This approach aligns with practices that emphasise open form, indeterminacy, and performer-system interaction, where the score is understood not as a fixed artefact but as a set of conditions that enable musical action [6].

This paper contributes to these discussions by proposing the concept of living audio scores, articulated through the design of Nada Sfera Playground, a web-based live

soundscape installation¹. In this system, environmental audio streams are continuously transformed through real-time processes such as granular synthesis, spatial direction, sound diffusion, and dynamic control of acoustic intensity. Each realisation of the work is shaped by interaction and listening, resulting in performances that are recognisable in structure yet never identical in detail. The sonic environment is therefore inherently variable, non-linear, and contingent, resisting fixed representation.

In response to this variability, Nada Sfera Playground introduces a listening-based, non-visual-focused notational framework that does not rely on visual score representation as its primary mode of instruction. Instead, living audio scores operate as perceptual roadmaps, articulating macro-structural trajectories—such as entry, transition, and exit—while allowing micro-level sonic behaviour to emerge through real-time interaction. These scores guide musical form through attention, spatial awareness, and sensitivity to texture, supporting coherent performance without prescribing specific sonic outcomes. Where visual or symbolic elements are present, they function only as secondary orientation cues and do not constitute a visual score.

By framing notation as a process that unfolds through listening rather than reading, this study situates Nada Sfera Playground within contemporary discourse on sound as score, non-visual-focused notation systems, and ecological listening, arguing that living audio scores offer a viable alternative to reading-oriented score paradigms in web-based and installation-based sonic practices.

2. SYSTEM OVERVIEW- NADA SFERA PLAYGROUND

Nada Sfera Playground forms part of the broader Nada Sfera ecosystem, a modular soundscape streaming and interaction framework developed as part of the author's PhD portfolio project *Seed of Life (SoL)*. Established in 2020 as a self-funded pilot project, Nada Sfera comprises three interrelated components: Nada Sfera Binaural Streamer, Nada Sfera Analyser and Nada Sfera Playground. The project has been supported through research and artistic collaboration with Locus Sonus and the Bristol Interaction Group, and has been previously disseminated through the Electroacoustic Music Studies Network Conference (EMS) 2021 [7]. Within this ecosystem, Nada Sfera Playground functions as the performative and notational layer, transforming live environmental sound streams into a listening-driven 'musical' environment.

Nada Sfera Playground is implemented as a web-based interactive sound-art installation utilising the Web Audio API for real-time audio processing, spatialisation, and interaction. Six environmental sound streams are delivered using an internet radio streaming mode based on Icecast, allowing continuous, low-latency transmission of live

binaural soundscapes from remote locations. The system is accessed through a standard web browser, allowing deployment across installation, rehearsal, and performance contexts without dedicated software installation. This platform choice supports accessibility, scalability, and distributed listening while maintaining sufficiently low latency for real-time sonic manipulation.

Environmental audio streams originating from Nada Sfera Binaural Streamer devices are routed into a modular audio graph within the Web audio API framework. Each stream is treated as an independent audio node, enabling individuals control over level, spatial position, diffusion, and granular transformation. The system architecture is deliberately designed to privilege sonic perception over visual feedback; while a minimal interface provides access to control parameters, 'musical' change is perceived primarily through listening. In this way, interaction within Nada Sfera Playground is oriented toward attentional and perceptual engagement rather than screen-based manipulation.

The title "Nada Sfera" reflects both sonic and spatial conceptual grounding. "Nada" refers to sound as a fundamental vibrational principle within South and Southeast Asian musical thought, while "Sfera" evokes spatial and environmental dimensionality. Together, the term situates the project within a framework where sound is understood as both material and space, aligning with the system's emphasis on immersive listening and distributed sonic environments.

2.1. Interaction, Variability, Indeterminacy, and Notational Implications

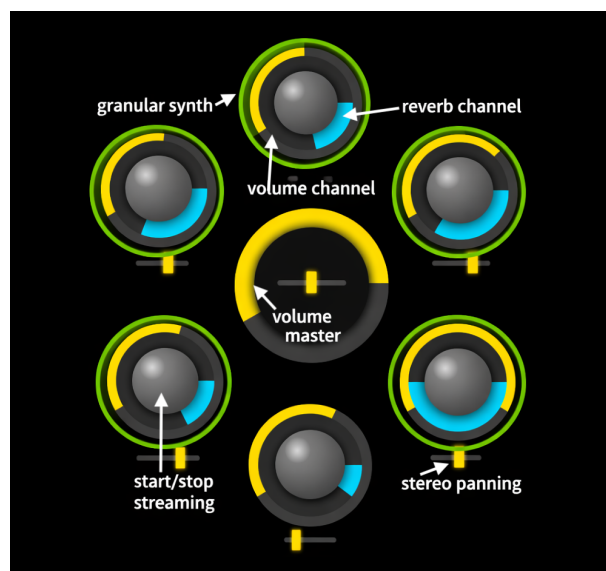


Figure 1. Nada Sfera Playground with improved user interaction (UI) design.

Within Nada Sfera Playground, participants interact with multiple real-time soundscape streams through a set of

¹ Nada Sfera Playground (prototype version) is available online: https://dac.uitm.edu.my/natural_soundscape/NadaSferaPlayground.html. The system is currently under ongoing development and may occasionally be unavailable due to maintenance. As a prototype, certain features, particularly panning and granular synthesis with an updated web

interface layout (see Figure 1), remain under active testing and may not be fully stable or consistently reflected in the online version.

continuous sonic parameters. These include channel activation and level control, spatial direction (panning) at both channel and master levels, sound diffusion articulated through spatial movement and reverberant depth, and real-time granular synthesis applied independently to each stream. Together, these parameters allow sound to transition fluidly between continuous environmental textures and fragmented micro-events, supporting attentive, exploratory listening (see Figure 1).

Importantly, the system does not impose linear progression or fixed ‘musical’ outcomes. Sound streams continue to evolve regardless of user input, positioning participants within an ongoing sonic ecology rather than a reactive instrument. Musical form emerges through sustained interaction, perceptual sensitivity, and the temporal behaviour of the system itself.

A defining characteristic of Nada Sfera Playground is its non-identical repeatability. Variability arises from multiple interacting factors: the contingent nature of environmental sound, the probabilistic behaviour of granular processes, and the interpretative choices of participants. Despite this openness, coherence is maintained through constraints embedded in the system design, such as bounded parameter ranges and gradual response curves.

These conditions establish a performative environment in which sound itself functions as an instructive medium. Rather than encoding musical content symbolically, the system generates a continuously unfolding sonic field that performers and visitors navigate through listening. This framing positions Nada Sfera Playground as a notational environment, preparing the ground for the introduction of living audio scores as perceptual guides that articulate structure without prescribing specific sonic results.

Although the available controls, such as level, panning, diffusion, and granulation, resemble those of a conventional mixing interface, their function within Nada Sfera Playground is fundamentally recontextualised. Rather than optimising balance or clarity, parameter adjustments are guided by perceptual listening states articulated through the score cards. The system therefore shifts from a corrective mixing paradigm to a performative and interpretative one, in which sonic relationships are shaped through attention, spatial awareness, and temporal sensitivity.

A recorded demonstration of a performance realisation of Nada Sfera Playground, including the use of audio score cards, is available at: https://drive.google.com/drive/folders/1BPuFHkree8Mdxn_AJdXZncPQzIAGLIP?usp=sharing

3. LIVING AUDIO SCORE IN NADA SFERA PLAYGROUND VIA AUDIO SCORE CARDS

3.1. Living Audio Scores as Listening-Based Notation

In Nada Sfera Playground, notation is reconceived as a listening-based process rather than a visual or symbolic representation of sound. Musical form is not specified in advance nor encoded through fixed graphical instructions;

instead, it emerges through real-time interaction with a continuously evolving real-time live soundscape. Within this context, the term living audio score refers to a notational framework that exists as an active sonic condition, unfolding through temporal, spatial, and textural change.

The score does not prescribe specific events; instead, it establishes conditions for musical action, aligning with traditions of indeterminacy and open form at the micro-structural level, while framing performance within a perceptually recognisable large-scale arc. Sound itself becomes the primary carrier of notational information, and listening becomes the central interpretative act through which performers and participants navigate form, respond to sonic behaviour, and shape ‘musical’ outcomes. In this model, notation functions as a facilitative process rather than an instructional artefact, supporting coherence without fixing outcome.

3.2. Manual-Pull Audio Score Cards

Navigation within the living audio score is supported through a system of audio score cards [8]. In the current implementation, these cards exist as external performative artefacts rather than embedded interface elements. They may be presented as printed cards or external digital prompts (e.g., screen-based decks), and are interpreted by performers alongside the web interface. Each card articulates a listening state rather than a sequence of actions or a time-based cue. Cards are activated exclusively through a “manual pull” mechanism, meaning that transitions are initiated by the performer based on listening judgement rather than automated timing or system triggers. The term “manual pull” therefore refers to a perceptual decision to shift listening state, not a software-driven event.

This approach explicitly rejects automated progression and elapsed-time logic. The duration of each card remains indeterminate and varies according to sonic behaviour, performer attention, and environmental contingency. As a result, the score remains reactive to sound rather than directive over it, preserving performer agency and foregrounding listening as a performative skill rather than reading as an instructional necessity.

The audio score cards are organised into three primary decks corresponding to a strophic arc-type formal trajectory: beginning, middle, and ending. This structure provides a recognisable large-scale form while avoiding linear timelines or teleological progression. The beginning deck establishes orientation, introducing spatial awareness and a shared listening posture. The middle deck functions as the core performative field, where density, texture, and spatial motion are explored; cards in this deck may be drawn repeatedly and in variable order. The ending deck supports dissolution and release, guiding the gradual withdrawal of sonic activity and attentional focus.

Movement between decks is determined solely through listening judgement rather than temporal markers. Performers may remain within the middle deck for an

indeterminate duration before initiating closure through the ending deck. In this way, the strophic arc provides perceptual clarity while remaining temporally open, allowing form to be realised through experience rather than prescription.

Cards are activated exclusively through a manual pull mechanism, meaning that transitions are initiated by the performer based on listening judgement rather than automated timing or system triggers. The term “manual pull” therefore refers to a perceptual decision to shift listening state, not a software-driven event. Each card articulates a listening intention that is realised through interaction with the system’s available parameters, such as channel level, spatial positioning, diffusion, and granular processing, establishing a consistent relationship between perceptual guidance and sonic outcome without enforcing fixed mappings.

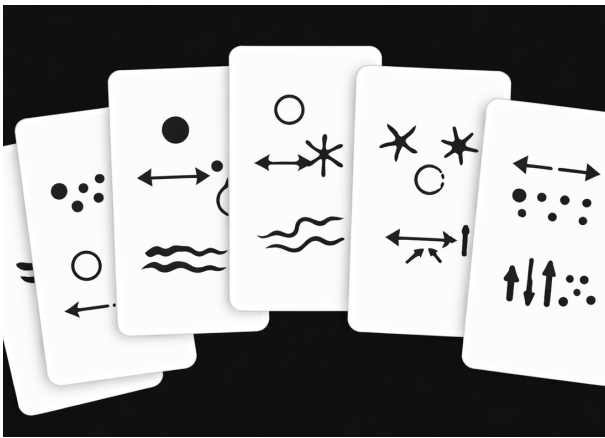


Figure 2. Example of external manual-pull audio score cards used in performance.

3.3. Performance Settings

Nada Sfera Playground supports both solo performance and distributed web-based ensemble realisations. In solo mode, a single performer operates within one web browser, navigating the three-deck card system while shaping the soundscape through listening-driven interaction. This approach relates to emerging browser-based and distributed score systems in which a shared state coordinates performance without prescribing identical actions, extending existing work in networked notation and collaborative digital performance environments [9, 10].

In distributed mode, multiple performers engage simultaneously from separate web browsers, forming a web-based installation ensemble. In this configuration, a shared card state is synchronised across participants, while sonic interaction remains local and interpretative. One participant assumes a scorekeeper role, responsible solely for manual card draws and transitions between the beginning, middle, and ending decks. Other participants respond to the active card by shaping sonic parameters, such as level, spatialisation, diffusion and granulation, through attentive listening.

This separation between a shared notational state and local sonic interpretation aligns with established approaches to network the browser-based score systems, where coordination is achieved without enforcing identical actions [9, 10]. The scorekeeper curates the evolving listening content rather than controlling sound, enabling collective musical behaviour to emerge through distributed listening.

Conceptually, the audio score cards share affinities with indeterminate and instruction-based practices, particularly those associated with Cage’s framing of compositional conditions and Fluxus event scores [11]. Live event scores, the card function as proposals rather than commands, specifying conditions under which musical action may occur rather than prescribing concrete outcomes.

However, unlike many object- or action-based Fluxus scores, Nada Sfera Playground situates instruction directly within an evolving sonic environment. The manual-pull mechanism further aligns the system with active listening and improvisational practice, particularly as articulated in Oliveros’ Deep Listening, where musical decisions arise from sustained attentional engagement with sound [2]. Improvisation here is understood not as gestural virtuosity, but as perceptual responsiveness to sonic change.

3.4. Audio Score Card Perceptual Symbols, Cognition and Interaction

Each audio score card combines short textual prompts with a limited set of perceptual symbols. These symbols do not encode musical parameters nor function as a visual score. Instead, they serve as orientation cues, reinforcing perceptual categories such as presence, distance, motion, and textural state.

The symbol system intentionally employs high-contrast, monochrome forms to minimise visual effort and reduce cognitive load. Research in visual cognition and icon legibility demonstrates that luminance contrast significantly affects recognition speed and response accuracy, particularly in attention-divided contexts [12, 13]. By avoiding colour-dependent encoding, the symbol design improves accessibility for users within colour-vision deficiencies and supports rapid orientation without drawing attention away from listening [14].

This design rationale parallels principles observed in early perceptual development, where which-constrast visual stimuli facilitate rapid pattern recognition. In Nada Sfera Playground, reducing visual complexity allows cognitive resources to remain focused on auditory perception and decision-making rather than visual decoding.

While the audio score cards operate at a perceptual and cognitive level, interaction within Nada Sfera Playground is realised through technical controls such as channel level, spatial positioning, sound diffusion and granular synthesis. The relationship between symbolic cues, perceptual intention, and possible technical approaches is exemplified in Table 1.

Symbols	Perceptual Meaning	Listening Intention	Possible Technical Approaches
●	Presence / Foreground	Bring one sound stream into perceptual focus	Increase relative channel level; reduce reverb send; narrow spatial spread
○	Distance / Withdrawal	Allow sound to recede from attention	Lower channel level; more reverb; widen spatial diffusion; reduce granular density
↔	Lateral Drift	Encourage spatial motion without destination	Apply gradual modulated left-right panning; avoid fixed centre positions
⋯	Granular / Fragmented Texture	Attend to micro-events and discontinuity	Increase grain density; randomise grain playback position; shorten grain size
...	Sparse / Pointillistic	Focus on isolated sonic events	Reduce number of active channels; lower grain density; introduce silence between events
◎	Resonant / Distant Space	Attend to spatial depth and reverberation	Increase reverb time and wet signal; diffuse spatial position

Table 1. Nada Sfera Playground Live Audio Score Card performance notes.

Although Nada Sfera Playground embraces indeterminacy and environmental contingency, coherence is maintained

through the combined effect of the three-deck structure, the manual-pull mechanism, and the perceptual consistency of the audio score cards. Across performances, the system exhibits recognisable formal behaviour while allowing substantial micro-level variation arising from sound sources, granular processes, and performer interpretation.

In this way, living audio scores enable ‘musical’ form to remain both open and intelligible. Rather than fixing outcomes, the score supports an ongoing negotiation between sound, system, and listener, demonstrating how notation can function as a dynamic, listening-based practice within web-based and installation-oriented sonic environments.

4. CONCLUSIONS

This paper has introduced living audio scores through the design of Nada Sfera Playground, a web-based live soundscape installation that frames musical form through listening rather than fully visual representation. By combining manual-pull audio score cards with a three-deck strophic arc ‘musical’ form, the system establishes conditions for musical action, while allowing sonic detail to emerge through real-time interaction. The proposed framework supports both solo and distributed web-based ensemble performance, demonstrating how a shared notational state can coordinate collective listening without prescribing sonic outcomes. By positioning notation as a perceptual roadmap rather than a fixed score, Nada Sfera Playground contributes a listening-based alternative to reading-oriented score paradigms, aligned with contemporary sound-as-score practices in web-based and installation contexts.

Acknowledgments

The author gratefully acknowledges the support of Universiti Teknologi MARA (UiTM), Faculty of Music, and the UiTM Computer Music and Musical Acoustics Initiative Research Group (i-COMMA) for supporting this research and the author’s participation in the TENOR Conference 2026. This participation is undertaken in conjunction with the parallel artistic research project supported by the ASEA-UNINET Grant 2025/2025 (ASEA-40/2025), titled Echoes of Tradition: A Cross-Cultural Artistic Research on Folklore and Natural Soundscapes. The author also thanks the collaborating institutions and partners involved in the ongoing development of the Nada Sfera project.

5. REFERENCES

- [1] H. Westerkamp, *What’s in a soundwalk?* [Online]. Available: <https://www.urbanimaginaries.com/s/CO-MS-5218-UI-Soundwalk-Guidelines.pdf>
- [2] P. Oliveros, *Deep Listening: A Composer’s Sound Practice*. New York, NY, USA: iUniverse, 2005.

- [3] B. Truax, *Acoustic Communication*, 2nd ed. Westport, CT, USA: Ablex, 2001.
- [4] J. LaBelle, *Acoustic Territories: Sound Culture and Everyday Life*. New York, NY, USA: Continuum, 2010.
- [5] S. Emmerson, *Living Electronic Music*. Aldershot, UK: Ashgate, 2007.
- [6] C. Cardew, "Notation–Interpretation, etc.," *Tempo*, no. 58, pp. 21–33, 1961.
- [7] Ainolnaim Azizol, "SOUNDING THE UNHEARD: REVIVING MALAYSIA'S NATURAL-CULTURAL SOUNDSCAPE THROUGH LIVE ELECTROACOUSTIC NARRATIVES," *Jurnal Gendang Alam (GA)*, vol. 15, no. 2, Dec. 2025, doi: <https://doi.org/10.51200/ga.v15i2.7143>.
- [8] I. Clester and J. Freeman, "ScoreCard: Generative music programs as QR codes." Available: <https://ijc8.me/static/scorecard.pdf>
- [9] R. Gottfried and G. Hajdu, "DRAW SOCKET: A BROWSER BASED SYSTEM FOR NETWORKED SCORE DISPLAY." Accessed: Dec. 28, 2025. [Online]. Available: <https://www.tenor-conference.org/proceedings/2019/05Gottfried.pdf>
- [10] Luca Turchet and Martina De Cet, "A web-based distributed system for integrating mobile music in choral performance," *Personal and Ubiquitous Computing*, vol. 27, no. 5, pp. 1829–1842, Feb. 2023, doi: <https://doi.org/10.1007/s00779-023-01709-0>.
- [11] L. Dorin, ed., *Fluxus Performance Workbook*. [Online]. Available: <https://www.thing.net/~grist/ld/fluxusworkbook.pdf>
- [12] Y.-H. Ko, "The effects of luminance contrast, colour combinations, font, and search time on brand icon legibility," *Applied Ergonomics*, vol. 65, pp. 33–40, Nov. 2017, doi: <https://doi.org/10.1016/j.apergo.2017.05.015>.
- [13] C. Ming, Z. Wu, H. Gu, C. Chen, and Y. Zhang, "The effects of luminance contrast and color combination on icon cognitive performance," *Color Research & Application*, Aug. 2021, doi: <https://doi.org/10.1002/col.22720>.
- [14] E. Zhang et al., "Experimental Study on the Universal Design of Signage Size and Brightness Contrast for Low Vision Individuals," *Buildings*, vol. 14, no. 7, pp. 2063–2063, Jul. 2024, doi: <https://doi.org/10.3390/buildings14072063>.

TRYING TO KEEP UP: MANAGING COGNITIVE LOAD IN AUDIO SCORES

James Saunders

Open Scores Lab, Bath Spa University
j.saunders@bathspa.ac.uk

ABSTRACT

This paper examines how cognitive load theory can inform the design of audio scores that deliberately create unstable performance situations. Audio scores present temporal information that may be unknown in advance and which constantly disappears, creating unique cognitive demands compared to static visual notation. Drawing on cognitive load theory from instructional design, I identify how intrinsic, extraneous, and germane cognitive load—alongside specific effects including split-attention, modality, transient information, and task switching—shape performer responses to dense streams of verbal instruction.

Unlike instructional design, which aims to reduce extraneous cognitive load to optimize learning, my compositional approach deliberately manipulates these parameters to create productive difficulty. Through discussion of two recent pieces I consider how varying the density, simultaneity, and sequentiality of verbal cues generates distinct performance conditions. The discussion explores how player control over pacing fundamentally alters the way cognitive load manifests, shifting from individual overwhelm to distributed decision-making. For composers working with audio scores, this suggests practical strategies: manipulating verbal instruction parameters provides direct control over extraneous load while keeping individual task complexity low, and incorporating repeated instructional patterns helps players develop response schema.

1. INTRODUCTION

Audio scores present complex information in time, requiring users to manage their responses in an environment where this information may not be known in advance and may constantly disappear. The flow of information may not be within the control of the player, instead being managed by the speed and density of delivery of information as an external audio stream. Whereas in static visual scores the stream of notational information that is needed is mostly available in advance, with studies [1] showing the way a musician ‘only rarely

fixates a bar ahead of what is currently being played’ [1, p. 18], in audio scores this is not generally the case. This categorical difference presents an opportunity to play with expectation and player responses to produce a more reflexive performance situation. In particular, the cognitive load placed on players when faced with a dense stream of instructions creates contingent situations where a host of strategies driven by heuristics might be employed to encourage certain musical and behavioural responses. The scenarios might engage with notions of failure, empathy and pragmatism which have the potential to reference situations we experience in everyday life, making connections with our lived experience and framing this in a playful and reflective manner.

In this paper I focus on *cognitive load theory* to understand how scores might be designed to manage user responses effectively. I begin by surveying recent scholarship from the field, isolating key aspects that might translate to music composition before presenting two examples of my own work that deploy these ideas to create unstable performance scenarios.

2. COGNITIVE LOAD THEORY AND AUDIO SCORES

The delivery of instructional information in an audio score can be thought of, in part, as a learning and comprehension task. The information is presented in a particular way with respect to its density, speed of delivery, simultaneity and sequentiality [2]. The configuration of these elements creates the score’s different affordances. Users must be able to decode and understand this information in time in order to carry out the score’s instructions. As such we might understand more about the way audio scores can be structured for preferred outcomes by looking at theory related to information learning and instructional design, specifically *cognitive load theory* and its framing of multimodal information delivery.

2.1. Cognitive Load Theory

Cognitive load theory is one of the principal frameworks for instructional design, which ‘deals with learning and problem-solving difficulty that is artificial in that it can be manipulated by instructional design’ [3, p. 295]. It is useful in relation to audio scores because it simultaneously considers ‘the structure of information and the cognitive architecture that allows learners to process that information’ [4, p. 1], taking into account the

Copyright: © 2026 James Saunders. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

way in which information is presented to a user as well as the inherent nature of the information itself as variables that impact understanding.

2.1.1. Element Interactivity and Intrinsic Cognitive Load

Cognitive load is driven by *element interactivity* [3], [4], which determines the way material can be learned, either sequentially or simultaneously. Sweller notes that ‘if elements can be learned successively rather than simultaneously because they do not interact, intrinsic cognitive load will be low’ [3, p. 295]. For example, learning the name of a single person in a group has a low level of element interactivity as the person-name pair does not have any further reliance on other information. Although the task may be made harder by being required to memorize a lot of names, each of these is a single instance within a sequential set of tasks as the name of one person is not reliant on the name of another. In contrast, if simultaneous ‘interactions between many elements must be learned, then intrinsic cognitive load will be high’ [3, p. 295]. For example, learning the correct set of moves to solve the Rubik’s Cube has high element interactivity as learning the result of a single move, while useful as a basic operation, is not sufficient to solve the complete cube which relies on being able to connect complex sequences of moves to achieve the desired result.

Element interactivity is therefore a defining factor of *intrinsic cognitive load*, which is ‘concerned with the natural complexity of information that must be understood and material that must be learned’ [5, p. 124]. In essence, some material is hard to learn and creates a high cognitive load due to its complex element interactivity, correspondingly making higher demands on working memory capacity. For audio scores, which may require players to recall complex information in real time while undertaking other actions, element interactivity and intrinsic cognitive loads are key indicators of the difficulty of the task.

2.1.2. Extraneous Cognitive Load

In contrast, *extraneous cognitive load* is imposed on this material by the instructional design. In the previous person-name example, learning one pair at a time in isolation at a comfortable rate has a low extrinsic load as each instance can be reinforced separately as an aid to retention. Consider though if the learning method was in a group of 20 people where each person gives their name in sequence once with the same requirement to learn person-name pairs. This would be a much harder task given the way the learning situation is presented, despite the task of remembering person-name pairs being the same. The intrinsic load is low, but the extrinsic load is much higher due to the instructional design.

Crucially, extraneous load is something that can be mitigated by careful instructional design, which is why it is of significance to making audio scores. As Sweller suggests in his original paper on cognitive load theory, ‘extraneous cognitive load ... is artificial because it is

imposed by instructional methods’ while intrinsic cognitive load is something ‘over which instructors have no control’ [3, p. 307]. For audio scores, careful management of the instructional design is within the control of the score-maker, enabling aspects such as density, speed of delivery, simultaneity and sequentiality to impact appropriately on player responses.

2.1.3. Germane Cognitive Load

A third category, *germane cognitive load*, is also a significant and controllable factor in instructional design, but one which enhances learning through devoting working memory resources ‘to schema acquisition and automation’ [4, p. 2]. Recent developments in the definition of cognitive load theory have repositioned germane cognitive load as ‘the cognitive load required to learn.’ Specifically germane cognitive load ‘redistributes working memory resources from extraneous activities to activities directly relevant to learning by dealing with information intrinsic to the learning task’ [6, p. 264]. It helps learners focus on the intrinsic elements of the task by recalling and developing new schema to apply to the task. For example, when learning to solve the Rubik’s Cube a network of heuristic observation and move techniques can form a schema to speed up user response to recognizable and previously experienced types of configuration. Instructional design can therefore leverage germane cognitive load to improve learning in future tasks. In audio scores, this might be through the use of common formulations of instructions which help players recall previous responses and strategies in an efficient manner, rather than treating each as novel information to process.

2.1. Cognitive Load Effects

Beyond the foundational categories of intrinsic, extrinsic and germane cognitive load and their impact on working memory, cognitive load research has isolated several distinct effects that instructional design creates [6]. Three of these—*split-attention effect*, *modality effect* and *transient information effect*—are of particular relevance to audio score design.

2.1.1. Split-attention Effect and Modality Effect

When learners are ‘required to divide their attention among and mentally integrate multiple sources of information’ [7, p. 319] split-attention effect can occur. For example, when looking at a diagram with spatially separate text explanations, there is a higher cognitive load on working memory as the learner must cross-relate and integrate two sets of information and determine their relationship to understand the material. If the same textual information is presented on the diagram as annotation, split attention effect is reduced as all the information is in the same place. As a parallel, consider the use of performance note prefaces in scores needing a lot of explanation and the impact on performer understanding when learning a new piece.

Split-attention effect is more prevalent when all the information is presented in the same domain, such as visual information (image-text, or text-text). Mousavi, Low and Sweller showed that when learners ‘must split their attention between multiple sources of information that require mental integration, cognitive resources available for learning can be increased by presenting some of the verbal material in auditory rather than written form’ [7, p. 333]. This *modality effect* suggests that presenting information in multimodal visual-auditory channels improves learning given the ‘existence of separate working memory channels or processors for auditory (especially verbal) and visual (especially pictorial) information’ [7, p. 319]. For audio scores, presenting some material in a visual mode may therefore reduce the load on working memory and improve the effectiveness of player responses.

2.2.2. Transient Information Effect

Given the temporal nature of audio scores, the *transient information effect* is highly relevant as it is concerned with ‘information that is presented to learners but disappears after a few seconds, for example, in spoken text or in instructional video or animation’ [6, p. 277]. This contrasts with non-transient information, such as a written text with pictures where ‘all information is available to the learner at the same time and may be revisited when needed’ [6, p. 277]. Having to store this disappearing information in working memory correspondingly increases the cognitive load, making the task harder to accomplish. Strategies to mitigate this problem include *self-pacing* where learners can manage the temporal flow of information, or *segmentation* where temporal information is presented in blocks with pauses in between [6, p. 278]. For audio scores, transient information is a fundamental reality but giving players opportunities to control the speed and density of instructions, or presenting breaks for them to settle, may also improve response. This may be contextually sensitive, however. Sweller, Van Merriënboer and Paas cite research by Leahy and Sweller which suggests ‘short pieces of audio-visual information were more effective than visual information only (i.e. traditional modality effect), but longer pieces of audio-visual information were less effective than visual information only because of the abundance of transient information in the longer, auditory piece’ [6, p. 278].

2.2. Task switching

Cognitive load can also be heightened by *task switching*, where learners must switch between different task types in logical, predictable, or randomized ways depending on the context. Where a task sequence requires learners to change what they are doing an interruption can occur, such that ‘it costs more, typically measured in reaction time, to switch between task sets than to repeat a task of the same type’ [8, p. 666]. This finding has been termed *switch costs*. As Kiesel et al. suggest, in ‘a task-cuing paradigm

with unpredictable sequences ... the order of the tasks and thus the order of task switches and repetitions are random’ resulting in performance that is ‘typically worse in switch trials than in repetition trials, revealing robust switch costs also in the task-cuing paradigm’ [9, p. 852]. This may be particularly noticeable in generative scores where a novel sequence of cues is generated for each version of the piece. Interestingly, where there is similarity between cues which may still produce different results, such as the verbal cue ‘g’ referring to a pitch, a major or minor chord, or a fragment of music, ‘switching between task sets with common elements such as shared stimuli or responses also imposes greater costs’ [8, p. 666].

2.3. Emotion, stress and uncertainty

Beyond the inherent configuration of cognitive load in a set of tasks, the presentation context for audio scores also impacts on performance. A further emergent effect on the ability of a player to respond to cues is caused by emotion, stress and uncertainty, which Sweller, Van Merriënboer and Paas, citing Moran, suggest ‘may restrict the capacity of working memory by competing with task-relevant processes; thus, they increase cognitive load, hamper learning and decrease transfer’ [6, p. 285]. In emotionally heightened and stressful states, such as performing complex tasks in front of an audience with a high and visible prospect of failure, it is likely that uncertainty will lead to errors. They note, however, that this is an integral part of some activities, such that learners must develop ‘professional competencies enabling them to perform professional tasks up to the standards, including the ability to deal with emotions, stress and uncertainty and to maintain overall wellbeing’ [6, p. 285]. Rehearsal and conditioning naturally mitigate this effect, but there remains an uncertainty where the flow of information to which players respond is outside their control.

In combination, the three categories of cognitive load theory and its emergent effects have significance for audio score design given the need for users to hold complex multimodal information in their working memory and respond in real time. To evidence this I present examples from two of my pieces where management of cognitive load has been part of the compositional design.

3. WORKING WITH COGNITIVE LOAD

In my recent work I have been exploring different ways to manage cognitive load. In instructional theory, much of the narrative is around clarifying processes to remove extrinsic cognitive load so that learners can focus on the salient features of the task and improve learning. My interest partly aligns with this, at least regarding understanding the task, but is more concerned with the way cognitive load can create ambiguity, error and the

unexpected. In doing so, my hope is that the experience of the player undergoing the task creates an empathic response for the audience, generally in relation to the way we experience overload in everyday life.

To do this, my recent pieces explore group decision-making and individual player agency in networks of interaction, thinking about how power structures emerge from groups working on a joint task. These pieces present players with many conflicting choices to create unstable dynamic situations that require constant attention. Individual preferences can be manifested but are always mitigated by the decisions of others in the group. These decisions are signalled orally, with players or artificial voices giving spoken cues that are also audible to the audience, making the decisions and their consequences explicit. The dynamically-produced audio score that emerges requires players to process different types of cue to complete the tasks as best they can, which can create different levels of cognitive load.

3.1. just trying to do what I'm told and keep up and do the best that I can (2025)

In *just trying to do what I'm told and keep up and do the best that I can* (2025) [10], a pianist responds to cues given by artificial voices, controlled by a bespoke app, attempting to follow all the instructions as they are given, as well as interacting verbally in conversation with an AI voice. In places the density of instructions may be too great to manage, but the aim is to try to follow as many cues as possible and adopt strategies to cope where this becomes difficult. The intrinsic cognitive load is relatively low as most of the cues are simple pairings with a required action. These include playing specific pitches or chords, cueing named samples, and using sound-making objects. When presented individually as discrete actions the task is relatively trivial, but the extrinsic cognitive load created by the design of the piece affects the level of difficulty due to the split-attention effect, modality effect, and transient information effect.

The deployment of cues can be varied when a version of the piece is set up, and typically this might involve a tutorial approach, where a pair of categories, such as major and minor chords, are used in isolation before adding in more categories. This helps the pianist, and also the audience, understand what is happening. Each category is assigned a different voice which helps differentiate between the categories, while also emphasizing the split attention of different cues operating within the same mode. Mousavi, Low and Sweller report on a 1972 experiment by Allport, Antonis and Reynolds which found that 'people could attend to and repeat (shadow) continuous auditory speech while processing unrelated visual scenes or sightreading piano music' [7] which suggests that the modality effect is helpful for audio scores that instigate instrumental performance. In contrast, however, the same experiment evidences the problem of receiving multiple verbal instructions, noting that 'if participants were presented with a series of words to remember, recall was worse if the words were

presented in auditory rather than visual mode...Hearing words while listening to speech interfered with memory of those words more than did reading them.' [7, p. 321] While not directly analogous with the situation the pianist faces in my piece, it may be significant in increasing the extraneous cognitive load.

As this is an audio score, transient information effect is also apparent as the cues are delivered in time requiring the player to store new information while simultaneously acting on previous information. The flow is outside the player's control so the cognitive load cannot be mitigated by self-pacing or segmentation, although this may be a feature of the score design. The transient information effect therefore impacts on player efficacy, especially in sections where there is a very dense, overlaid sequence of cues given across multiple categories.

The other significant layer in the piece—holding conversations with the AI—further taxes the player's working memory by introducing an additional activity to attend to. There is a distinct difference involved in responding mechanically to simple word-action cue pairs and holding a free-form conversation with a fairly unresponsive AI persona. The difficulty here may be explained by the cost associated with task switching, especially as engaging in the creative thinking necessary in conversation is set against listening and responding to instructions and performing a range of disparate physical performance actions. The player has some control over what they choose to do—they may decide or forget to talk to the AI for instance—but this too may be a necessary way to manage the cognitive load when there is too much to do.

3.2. it is really hard to focus when there are too many things to do (2023)

Managing tasks in this way occurs slightly differently in pieces where players do have more control over the pacing of cues. In my ensemble piece *it is really hard to focus when there are too many things to do* (2023) [11] players are presented with a collection of tasks which involve completing different series of actions, including playing melodic patterns, sequences of object sounds and samples. Individual players may cue these sequences, affecting the behaviour of the group. Once a sequence has been started, players must complete it while trying to manage the completion of any unfinished sequences. Simultaneously, other instructions modify the way they complete these tasks, necessitating decisions as to how best to proceed to do the best they can.

Here the instructions are in part delivered by other players, so are in effect extraneous for each individual, but players are also able to deliver their own instructions so the transient information effect can be partially reduced through self-pacing. In this piece the instructions are also delivered live by human performers who may respond to the current situation, which contrasts with the piano piece where the instructions are regulated in a controlled random manner by the app. This may introduce

segmentation as the ensemble tacitly agrees to pause instruction giving to reduce the cognitive load (or for other reasons, such as satisfaction with the current musical material).

Other noted effects are apparent. Split-attention effect and task switching are highly relevant as players playfully cue new actions and musical material. In parallel a rising and falling numerical count determines when the piece will finish, with players shouting out numbers one higher or lower than the previous number, and the piece ends when the count gets to zero. Consequently, players are required to track multiple cues related to different types of action. Frequently these will interrupt each other, such as beginning a new pattern while finishing playing a sequence of samples. The high task switching costs create additional cognitive load as players need to make executive decisions about how to resolve incompatible instructions where they are required to do two or more things at once.

4. CONCLUSION

The discussion of these two pieces demonstrates how cognitive load theory provides a productive framework for understanding and designing audio scores that deliberately create unstable performance situations. In *just trying to do what I'm told and keep up and do the best that I can*, the combination of split-attention effect, transient information, and task switching costs creates moments where the pianist must selectively attend to some cues while necessarily abandoning others, developing real-time coping strategies such as forgetting to engage with the AI conversation when other demands become overwhelming. In *it is really hard to focus when there are too many things to do*, the ensemble's collective agency introduces different dynamics: players can self-pace and segment instruction delivery through tacit coordination, partially mitigating transient information effect, though this comes at the cost of heightened task switching as they navigate multiple concurrent sequences and the numerical countdown. The contrast between these pieces suggests that player control over pacing fundamentally alters how cognitive load manifests, shifting from individual overwhelm in the solo piece to distributed decision-making in the ensemble context.

This application of cognitive load theory inverts its conventional purpose. Where instructional design typically aims to reduce extraneous cognitive load to optimize learning, my compositional approach deliberately manipulates these same parameters to create productive difficulty. This inversion is artistically meaningful rather than simply perverse because it makes visible the processes of coping, failure, and adaptation that characterize contemporary experience more broadly. The audience hears not only the musical results but also the explicit verbal cues and performer responses, creating what I hope becomes an empathic engagement with the experience of managing competing demands under pressure. The ethical dimension of deliberately imposing

overload on performers is mitigated by transparency—players know what they are entering, the difficulty is part of the artistic proposition, and the piece's title and structure make explicit that 'doing the best that I can' is the goal, not flawless execution.

For composers working with audio scores, this suggests several practical considerations. Manipulating the density, simultaneity, and sequentiality of verbal instructions provides direct control over extraneous cognitive load, while the intrinsic complexity of individual tasks can remain relatively low. Incorporating germane cognitive load through repeated instructional patterns helps players develop schema for response, making the overall task more navigable even when density increases. The modality effect suggests that combining visual and auditory information may reduce overall load, though in some contexts—as with the piano piece—maintaining everything in the auditory domain may be preferable for aesthetic reasons. Most significantly, decisions about player agency over pacing emerge as crucial: giving performers control fundamentally changes the experience from endurance to negotiation.

Cognitive load theory and its resultant effects present a useful lens through which to view audio score design. While some of these observations about audio scores are perhaps unsurprising as dealing with lots of dense simultaneous spoken instruction is obviously hard, the nuance presented by the theory may prove to be useful in structuring some elements of a score. Given that this field serves as a theoretical basis for instructional design, drawing on its strategies for managing cognitive load in learning tasks is in many cases directly transferable to scoring. In particular, considering how extrinsic cognitive load can be balanced to create required performance outcomes and germane cognitive load can help develop schema to support players' ability to complete tasks effectively, a better understanding of how players navigate this form of notation might emerge.

5. REFERENCES

- [1] J. Perra *et al.*, "Review on eye-hand span in sight-reading of music," *Journal of Eye Movement Research*, vol. 14, no. 4, 2021.
- [2] J. Saunders, "Perfect information: scores as self-evident processes," in *Ninth International Conference on Technologies for Music Notation and Representation*, 2024, pp. 53–57. Available: <https://www.tenor-conference.org/proceedings.html>.
- [3] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learning and Instruction*, vol. 4, no. 4, pp. 295–312, 1994.
- [4] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory and instructional design: Recent developments," *Educational Psychologist*, vol. 38, no. 1, pp. 1–4, 2003.
- [5] J. Sweller, "Element interactivity and intrinsic, extraneous, and germane cognitive load," *Educational*

Psychology Review, vol. 22, no. 2, pp. 123–138, 2010.

[6] J. Sweller, J. J. G. Van Merriënboer, and F. Paas, “Cognitive architecture and instructional design: 20 years later,” *Educational Psychology Review*, vol. 31, no. 2, pp. 261–292, 2019.

[7] S. Y. Mousavi, R. Low, and J. Sweller, “Reducing cognitive load by mixing auditory and visual presentation modes,” *Journal of Educational Psychology*, vol. 87, no. 2, pp. 319–334, 1995.

[8] M. Chis et al., “The cognitive load – productivity tradeoff in task switching,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 67, no. 1, 2023, pp. 666–671.

[9] A. Kiesel et al., “Control and interference in task switching—a review,” *Psychological Bulletin*, vol. 136, no. 5, pp. 849–874, 2010.

[10] J. Saunders, “just trying to do what I’m told and keep up and do the best that I can” [musical score], 2025. Available: <https://researchspace.bathspa.ac.uk/17643>.

[11] J. Saunders, “it is really hard to focus when there are too many things to do” [musical score], 2023. Available: <https://researchspace.bathspa.ac.uk/15305>.

EXPERIMENT DESIGN FOR PSYCHOACOUSTIC ANALYSES OF AUDIO-SCORE COMPOSITIONS

Katharina Pollack

Acoustics Research Institute
Academy of Sciences, Vienna
katharina.pollack@oeaw.ac.at

Piotr Majdak

Acoustics Research Institute
Academy of Sciences, Vienna
piotr.majdak@oeaw.ac.at

Barbara Maria Neu

University of Music and
Performing Arts, Vienna
neu@mdw.ac.at

ABSTRACT

Psychoacoustics and the perception of sound play a crucial role in the process of composing a musical piece. The interpretation by the musician, despite a-priori knowledge about the musical piece, and, if applicable, the interaction between the musician and the composer in a live scenario, both heavily depend on their auditory systems and auditory environment. To better understand the underlying psychoacoustic processes, we conducted an experiment using isolated parts of audio-score recordings of a clarinet. In the experiment, we focused on perceptive parameters involved in the musical piece. We relate our findings to existing psychoacoustic models, considering the limits of the models' auditory stages and the possibility of incorporating the models in the process of audio-score compositions. To this end, we give an outlook on what a psychoacoustic analysis of audio-score compositions could look like.

1. INTRODUCTION

A score contains information the composer wants to convey to the musician. It can comprise instructions regarding pitch, dynamics, and tempo, even instructions open to interpretation and possibly relying on mimesis, e.g., instructions regarding timbre. There are various interpretations of how an audio score is defined. For example, Bell distinguishes between reactive and rehearsed audio scores [1]. Bhagwati distinguishes between situative and fixed, and five conveyance subcategories [2]. He also states that “Musical precision is conveyed most effectively through music itself.” in the context of musical pedagogy. Similar to situative and fixed audio scores, Sdarulig et al. distinguish between online and offline audio scores [3]. And Schimana explores the concept of audio scores based on acoustic memory [4].

An audio score is not just a different visual representation of sound, e.g., a spectrogramme or a sketch of the score beyond classical notation. From a basic research perspective, a visual score represents the score in space, and an audio score represents the score in time. The score itself is transferred from the composer(s) to the musician(s) via at least

one acoustic transfer path. There can be multiple acoustic transfer paths, sometimes on purpose and sometimes unpredictable or even unwanted ones, because, e.g., the score can be conveyed by a loudspeaker to a musician, but the acoustic properties of a room and the number of instruments and loudspeaker playing simultaneously add another acoustic transfer path.

Using a computational psychoacoustic model to simulate a virtual listener can be advantageous for a number of reasons, one of which is to estimate the behaviour of a person prior to conducting an experiment. However, all psychoacoustic models are an approximation of reality and can focus on different stages of the auditory processing chain. Additionally, the virtual listener is not represented in the same way in all psychoacoustic models. One way would be to model the listener as an agent in a perception-action cycle [5]. In this perception-action cycle, the agent, in this case a musician, is interpreted as agent whose sensors receive signals from the environment, e.g., auditory, visual, or somatic signals. These sensations update the internal beliefs of the agent and influence movement, e.g., eye, head, or limbs, which in turn influence the environment and, thus, close the perception-action cycle.

In order to apply psychoacoustic models as composition tools, they have to be evaluated on synthetic sounds. In this study, we present a pilot study to help design an experiment setup for psychoacoustic analysis of audio-score compositions.

2. METHODS

In order to analyse why the musician played what they heard, we set up an experiment, in which we examined simple audio scores played by a single musician repeating isolated musical sequences. A video recording of the musician commenting on each take was made immediately after the recorded sequence in order to post-hoc identify conscious decisions.

2.1 Signals

Sequences were implemented in MaxMSP (8.6.3) on a laptop (MacOSX Sequoia 15.5). Using a sound pressure level meter, it was ensured that sound pressure levels at the ear of the musician would not exceed 85 dB(A). Three different tone combinations were created: [d', d#', f'], [a', h', c''], and [g#'', a'', b'']. Each combination was chosen to be within a critical band [6, 7], hence, the combinations

were called CB1, CB2, and CB3, respectively. For each of the combinations, four different sequences were created: *Long tones*, in which every tone lasted a few seconds, *long to short tones*, in which the duration of tones was shortened in an arhythmical way, *short tones*, in which every tone lasted less than one second, and *long tones with noise*, in which every tone lasted a few seconds but was accompanied with noise. For each of the sequences, the tones were played in a partially overlapping, partially alternating fashion with varying attack, decay, sustain, and release times. For the experiment, the following thirteen sequences were recorded:

1. CB1 long tones
2. CB2 long tones
3. CB3 long tones
4. CB3 long tones
5. CB1 long tones
6. CB1+CB2+CB3 long tones
7. CB1 long to short tones
8. CB2 long to short tones
9. CB3 long to short tones
10. CB1+CB2+CB3 short tones
11. CB1 long with noise
12. CB2 long with noise
13. CB3 long with noise

Basically, the sound was moving from one pitch to the next with varying speed, duration, and overlap time.

2.2 Participant and apparatus

The signal flow in the experiment was as follows: The signal generated by the computer of the composer was routed to an interface (RME Fireface UCXII 24110382) and played to the musician – the participant – over a loudspeaker (Genelec 8020B active monitor) placed 25 cm away from the musician’s right ear. A microphone was placed just above the musician’s ear and held in place by clipping it to an elastic headband. A second microphone was placed in the room in order to record the musician’s thoughts after the recordings, alongside the video recording of the camera. On the clarinet, one microphone was inserted by 5 cm into the cone of the instrument, and a gooseneck microphone 5-cm away from the cone. These two microphones were placed such that different timbre features may be extracted for further processing.

The room (3.5 x 5.4 x 3.2 m, width x length x height) was acoustically treated with Basotect plates on the walls. Using a sound-pressure level metre, a noise floor of 27 dB(A) was measured. In total, four people were in the room: audio recording engineer, camera operator, composer, musician. Towards one side of the room, the recording engineer

was placed behind a curtain. In the remaining part of the room, the composer, the musician, and the camera operator were placed in a triangular formation. The space between the camera operator and the composer was 1.5m, as was the space between the camera operator and the musician. The space between the musician and composer was 2.5 m, and the space between the musician to the wall was 80 cm. A loudspeaker (Genelec 8020B) was placed at the height and the distance of 25 cm from the musician’s ear with the musician sitting in a rigid position. During the recording, the distance between the musician’s ear and the loudspeaker changed marginally. In this scenario, no eye-contact during the recording was made between the composer and the musician in order to replicate a live-performance scenario as close as possible.

2.3 Procedure

Before the recordings, the instruction “play what you hear” was given by the composer to the musician. After each stimulus was recorded, the musician reported what they heard and what sound(s) they were trying to create. Sometimes, but not often, the composer would clarify on what part of the sound the musician should focus on and try to reproduce using their instrument.

3. RESULTS AND DISCUSSION

The total duration of the recording was two hours and 23 minutes and an analysis of the complete recording is out of the scope for this article. Here, we analysed three short clips from three different sequences, lasting only a few seconds. The first clip was from sequence 1 (CB1 long tones) in which the three tones were played ordered [f’, d’, d#’]. Figure 1 shows the waveforms with normalised amplitude of the sequence in various stages of the processing flow.

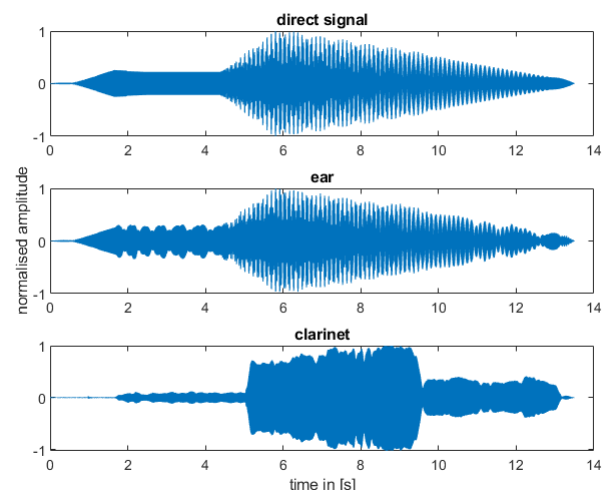


Figure 1. CB1, long tones. Top waveform is the direct signal, middle waveform is the signal arriving at the musician’s ipsilateral ear, bottom waveform is the signal measured 5 cm outside of the clarinet’s cone.

Direct signal was the output of the interface, the *ear* was

the microphone signal placed close to the musician's ipsilateral ear, and *clarinet* was the microphone signal placed outside the cone of the instrument. It can be seen that the direct signal from the interface is a synthetically generated one, containing linear amplitude changes. The amplitude of the ear signal is different than in the direct signal, because of reflections of the musician's head, shoulders, and ear. And in the clarinet signal, the musician started playing approximately 400 ms after the ear signal, a delay which consists of the musician's reaction time, their time to make a decision on what to play, and also the time to breathe in and start playing.

We also analysed the signals in the time-frequency domain. Figures 2, 3, and 4 show the spectrogrammes of the clip. These figures were created with mullaclab from the large time-frequency analysis toolbox (LTFAT) v2.6.0 [8, 9]. The window length was set to 4096 samples in order to increase the frequency resolution, other frame parameters remained at their default values. The beat can be seen clearly as well as the changing frequency (slow between f' and d' and fast between d' and $d\#'$).

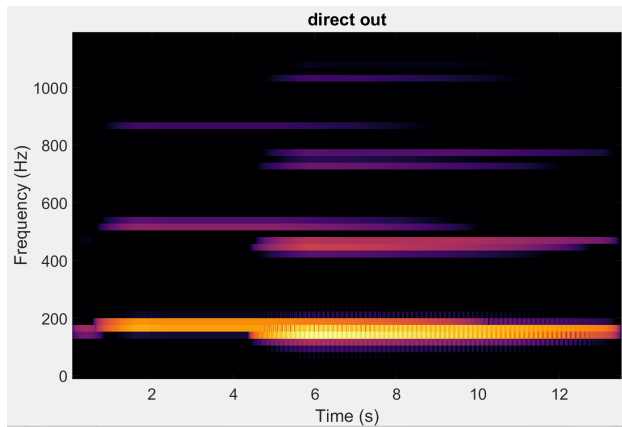


Figure 2. The signal *direct out* from clip 1 in the frequency domain.

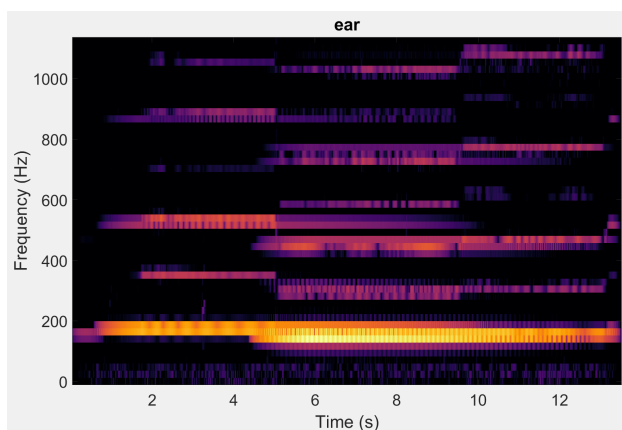


Figure 3. The signal *ear* from clip 1 in the frequency domain.

Changes in timbre can be seen between the direct-out and the clarinet signals (independent of the microphone position). In the *ear* signal, the beat started after two seconds

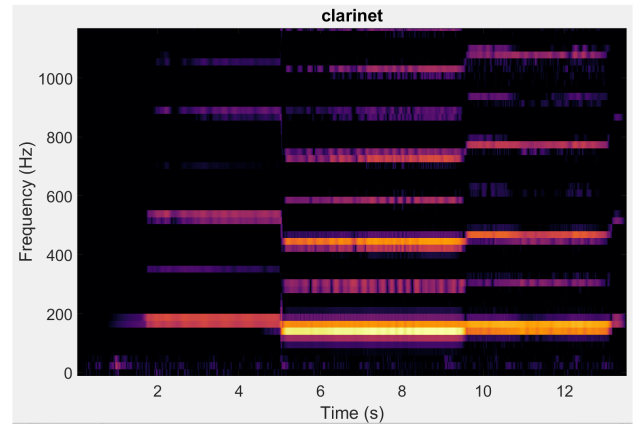


Figure 4. The signal *clarinet* from clip 1 in the frequency domain.

between f' and d' , and increased its frequency as the interval progressed to a minor second (d' and $d\#'$). The beat was not played by the musician, the musician only decided to switch notes once the next tone was louder.

Then we analysed two additional clips, one taken from sequence 7 (CB1 long to short) and the other from sequence 13 (CB3 long with noise). Figure 5 shows the clip from sequence 7. In the beginning of this clip, the musician played significantly shorter tones than what they heard, and it seems that later in the sequence, the amount of tones subsequently played were too much to play back the same way. Instead, the musician played only the tone (sometimes multiple tones) they could hear once they stopped playing.

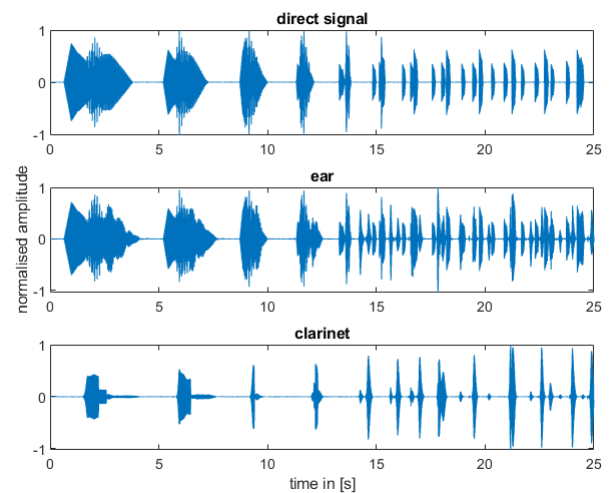


Figure 5. Clip from sequence 7 (CB1 long to short). Order of waveforms same as in Fig. 1.

Figure 6 shows the clip from sequence 13. In this clip, the direct-out signal contained noise underlying long tones. The clarinet signal did not contain noise, because apparently the musician focused on playing high-pitch tones quietly, a design choice by the composer which itself was already a challenge on this instrument. Because of the low volume of the speaker signal, the sound from the clarinet

was captured not only in the clarinet microphone but also in the ear microphone, which complicated the signal separation and, thus, the analysis.

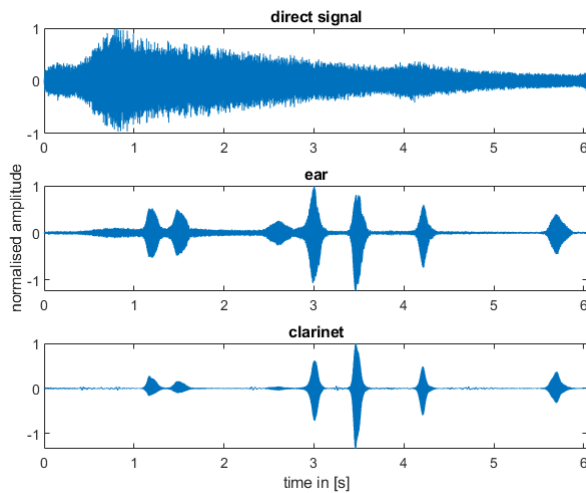


Figure 6. Clip from sequence 13 (CB3 long with noise). Order of waveforms same as in Fig. 1.

There are a number of psychoacoustic models that would be useful for compositions, for example, models covering temporal [12] or spectral masking [13, 14, 15, 11]. They could tell at what time or level differences two distinct sounds can be perceived.

With respect to spectral changes, there are two models that might be able to cover the prediction of whether or not a virtual listener could detect a change in timbre. One is a monaural model, i.e., not considering any binaural effects (e.g., location of a sound source) [15, 16, 17]. The other one is a binaural model [18].

One current limitation of these models is that they consider only static sounds. For example, a change can be detected in a sound that contains two clicks which are 40-ms delayed, but the model would not be able to handle a sound that contains more clicks with varying delays. Another known limitation of the perceptual similarity models is that they have a binary output whether or not a listener be able to detect a difference between the two sounds presented. However, there is no absolute threshold for change detection in timbre that is the same for each individual, and a probabilistic output, i.e., how likely is a listener able to detect a difference between two sounds, would make more sense in a realistic scenario.

4. CONCLUSIONS

In general, a new pitch introduced in the direct signal resulted in acoustic phenomena in the ear signal, such as, e.g., amplitude modulation. Exploiting such (psycho-)acoustic phenomena explaining whether a musician is able to hear a difference in audio scores can help composers to experiment with the auditory system of the musicians. In this contribution, we proposed an experiment setup that was piloted on one musician and instrument, and we covered

a first insight into the potentials of psychoacoustic analysis of audio scores. An application of said models in the context of audio scores seems to be required to better understand the interpretation of audio scores from the perspective of psychoacoustics.

Acknowledgments

This research was funded by the Austrian Science Fund (FWF) [grant DOI 10.55776/AR824]. The authors thank Elisabeth Schimana for providing the audio-score compositions.

5. REFERENCES

- [1] J. Bell, “Audio-scores, a resource for composition and computer-aided performance,” Ph.D. dissertation, Guildhall School of Music and Drama, 2017.
- [2] S. Bhagwati, “Elaborate audio scores: Concepts, affordances and tools,” in *Proceedings of the 4th International Conference on Technologies for Music Notation and Representation—TENOR*, vol. 18, 2018, pp. 24–32.
- [3] C. Sdraulig and C. Lortie, “Recent audio scores: Affordances and limitations,” in *Proceedings of the 5th International Conference on Technologies for Music Notation and Representation—TENOR*, vol. 19, 2019, pp. 38–45.
- [4] E. Schimana, “Sound as score: Live generated audio score and audio score based on acoustic memory,” in *Proceedings of the 5th International Conference on Technologies for Music Notation and Representation—TENOR*, 2019.
- [5] P. Majdak, R. Baumgartner, and C. Jenny, “Formation of three-dimensional auditory space,” in *The technology of binaural understanding*. Springer, 2020, pp. 115–149.
- [6] H. Fletcher and W. A. Munson, “Loudness, its definition, measurement and calculation,” *Bell System Technical Journal*, vol. 12, no. 4, pp. 377–430, 1933.
- [7] H. Fletcher, “Auditory patterns,” *Reviews of modern physics*, vol. 12, no. 1, p. 47, 1940.
- [8] P. L. Søndergaard, B. Torr sani, and P. Balazs, “The linear time frequency analysis toolbox,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 10, no. 04, p. 1250032, 2012.
- [9] Z. Pr  a, P. L. Søndergaard, N. Holighaus, C. Wiesmeyer, and P. Balazs, “The large time-frequency analysis toolbox 2.0,” in *International Symposium on Computer Music Multidisciplinary Research*. Springer, 2013, pp. 419–442.
- [10] A. Osses Vecchi, R. Garc a Le n, and A. Kohlrausch, “Modelling the sensation of fluctuation strength,” in *Proceedings of Meetings on Acoustics*, vol. 28, no. 1. Acoustical Society of America, 2016, p. 050005.

- [11] P. Majdak, C. Hollomey, and R. Baumgartner, “Amt 1. x: A toolbox for reproducible research in auditory modeling,” *Acta Acustica*, vol. 6, p. 19, 2022.
- [12] J. F. Culling, M. L. Hawley, and R. Y. Litovsky, “The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources,” *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 1057–1065, 2004.
- [13] T. Dau, D. Püschel, and A. Kohlrausch, “A quantitative model of the “effective” signal processing in the auditory system. i. model structure,” *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3615–3622, 1996.
- [14] —, “A quantitative model of the “effective” signal processing in the auditory system. ii. simulations and measurements,” *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3623–3631, 1996.
- [15] T. Dau, B. Kollmeier, and A. Kohlrausch, “Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers,” *The Journal of the Acoustical Society of America*, vol. 102, no. 5, pp. 2892–2905, 1997.
- [16] —, “Modeling auditory processing of amplitude modulation. ii. spectral and temporal integration,” *The Journal of the Acoustical Society of America*, vol. 102, no. 5, pp. 2906–2919, 1997.
- [17] A. Osses Vecchi and A. Kohlrausch, “Perceptual similarity between piano notes: Simulations with a template-based perception model,” *The Journal of the Acoustical Society of America*, vol. 149, no. 5, pp. 3534–3552, 2021.
- [18] T. McKenzie, C. Armstrong, L. Ward, D. T. Murphy, and G. Kearney, “Predicting the colouration between binaural signals,” *Applied Sciences*, vol. 12, no. 5, p. 2441, 2022.

COMPOSING THE EARS OF THE MACHINE: AUDIO SCORES, SELECTIVE LISTENING, AND CONCATENATIVE AI IMPROVISATION

Ming Yang

Royal Northern College of Music
(RNCM)

nicholasyang2016@gmail.com

ABSTRACT

This paper proposes an expanded notion of audio score for AI-mediated improvisation, grounded in psychoacoustic theory and practice-based composition. Building on Louis d’Heudière’s account of audio scoring with verbal instructions and field recordings, it shifts the focus toward sound without instruction as score: sound that guides performers through selective listening. Drawing on auditory scene analysis and timbre perception, the paper argues that sound can function as a score because listeners extract perceptual dimensions such as brightness, roughness, envelope, and tonal stability and treat them as cues for musical action. It then extends this framework to machine listening, treating corpus-based concatenative systems as computational analogues of selective, feature-based listening, in which descriptors, weights, and distance models define how the machine “hears” and which aspects of the input acquire prescriptive force. A case study of *Overlapping* for alto saxophone and live electronics examines how this idea is realized through two contrasting corpora, three designed saxophone sound types, and three microphones functioning as independent “ears,” creating a looping audio score shared by human and algorithmic agents. The paper concludes by discussing the limitations of descriptor-based listening—its partial view of the sound object—and argues for composing with multiple, parallel listening modes as a way of articulating richer audio scores between humans and machines.

1. INTRODUCTION

In Louis d’Heudière’s thesis, audio scoring is described as a practice in which sound functions as a medium of instruction, enabling performers to respond to recorded or spoken material rather than to symbolic notation [1]. Within his works, both field recordings and verbal instructions are treated as scores that guide performance: he characterises the field recording as a form of “stimulus” [1, p.

33] and allows both the recording and the verbalised instruction to appear in the concert situation as visible and audible agents [1, p. 33]. While d’Heudière explores the coexistence of verbal and sonic instructions, the present study shifts the focus toward sound without instruction as score—that is, toward the capacity of sound itself to establish communication between performers without the mediation of language. I take as a starting point her use of field recording as a guide to performance and extend it by asking what happens if we consider the influence of sound on performers’ cognition and attention as a general model of scoring. If field recordings can function as scores by directing human perception, then an analogous principle could apply wherever sound is selectively perceived and acted upon.

On this basis, I propose a further extension of the audio score concept into the domain of AI-mediated improvisation, where machine listening, and corpus-based concatenative systems (CBCS) participate in the same communicative field as human musicians. This move aligns with Pierre Alexandre Tremblay’s work on concatenative systems, which demonstrates that musical outcomes are shaped by the mapping of input descriptors to target descriptors and by their relative weights [2]. Loudness, spectral features, envelope, and other dimensions are extracted from the incoming sound and used to select or generate new material. Composition, in this view, resides not in predetermined notes but in the design of listening—in deciding which aspects of sound will be heard and acted upon. Following Tremblay, I understand machine listening as a process that extracts specific perceptual facets—brightness, roughness, gesture—and uses them to guide musical response.

The central question of this paper therefore becomes: how can audio scores operate when one of the listeners is an algorithmic system? In the proposed framework, both human improvisers and machines act through selective perception. Performers choose which elements of a sound to follow; the concatenative system does the same through weighted descriptors. The audio score is not only the sound presented to the algorithm; it is also the sound returned by the algorithm, which in turn influences the subsequent decisions of the performer. The score emerges from the multi-directional interaction between human improvisation and AI improvisation, understood as a continuous exchange of perceptual cues rather than as a one-way transmission of instructions.

2. LISTENING AS SELECTION: PSYCHOACOUSTICS AND THE CONDISIONT OF AUDIO SCORE

Selective listening during improvisation

As noted before, the fundamental question is: how can a sound, without verbal instruction, become a meaningful instruction for performers or systems? Sound can only function as a score if listeners do not hear all of it equally. Auditory perception is not a neutral recording of acoustic reality but a process of selective attention in which certain features become salient while others recede into the background. Work in auditory scene analysis has shown that the auditory system continuously groups and segregates components of complex sound mixtures into perceptually meaningful streams, based on features such as frequency content, temporal patterning, and spatial cues [3]. Psychoacoustic and timbre research further demonstrates that dimensions such as spectral brightness, roughness, temporal envelope, and tonal stability play different roles in how sounds are recognized, grouped, and remembered [4], [5]. A sharp attack can dominate the identification of an event, a noisy spectrum can attract attention, and a stable pitch can anchor perception, even when these attributes coexist within the same sound.

This selective nature of hearing is what allows sound to operate as instruction. When a performer hears a sound in an improvisational context, they do not respond to its total acoustic content but to those elements that stand out as perceptually or musically significant. A long-sustained tone may invite continuation, a burst of noise may call for contrast, and a bright timbre may suggest further spectral activity. In this way, sound becomes a field of affordances rather than a fixed object, guiding action through perceptual emphasis rather than symbolic encoding.

Psychoacoustic studies of timbre perception provide a formal basis for this idea. Timbre is not a single attribute but a multidimensional perceptual space structured by parameters such as brightness, roughness, attack character, and spectral flux [4]–[6]. Research on psychoacoustic sonification and auditory displays similarly identifies pitch, loudness, brightness, and roughness as separable perceptual dimensions that can be manipulated independently and interpreted reliably by listeners [7]. Listeners weigh these dimensions differently depending on context, task, and attention, resulting in variable judgments of similarity, salience, and grouping. Two sounds that are acoustically complex may be perceived as similar if they share a dominant dimension, or as different if they diverge along a salient axis.

Psychoacoustic Evidence for Perceptual Dimension Extraction

Having established that an audio score operates through the selective perceptual salience of sound, the question arises: what empirical evidence supports the idea that listeners extract particular dimensions from sound and use them to guide action? Psychoacoustics provides a scientific foundation for understanding how listeners attend to

different elements of sound. Decades of research in timbre and auditory perception have identified a set of core perceptual attributes—including loudness, roughness, brightness, tonalness, and temporal shape—those listeners treat as distinct sensations, even when these attributes arise from the same acoustic signal [5], [6]. In a series of experiments that aim to characterise perceptual dimensions of sound for auditory displays, researchers such as Tim Ziemer and colleagues have demonstrated that listeners can reliably extract such qualities from synthesized auditory stimuli and use them to perform cognitive and motor tasks [7], [8], [9].

In Ziemer's work on perceptual dimensions and psychoacoustic auditory displays, the focus is on designing mappings between data and sound such that listeners can extract intended information from specific perceptual features. In one set of studies, participants were presented with sounds that varied systematically along perceptually orthogonal dimensions—for example, changes in spectral centroid (related to brightness), modulation rate (related to perceived roughness), and pitch trajectory—and were asked to use these auditory cues to complete tasks such as judging category differences or navigating toward a target in a two-dimensional space [7], [8], [9]. Trajectory analyses indicate that users are able to analyse the sonification axis-by-axis and to integrate multiple sonified dimensions when approaching a target, showing that listeners can attend to and discriminate several auditory features simultaneously. The success of these experiments demonstrates that listeners do not treat complex sounds holistically or undifferentiated; rather, they hear perceptually salient dimensions and can isolate and act upon these dimensions when they carry meaningful information.

This line of research parallels classic psychoacoustic studies on timbre perception, which show that listeners do not perceive composite sounds as monolithic wholes but as configurations of distinct attributes. Multidimensional scaling of dissimilarity ratings between instrument tones reveals that changes in spectral centroid strongly influence perceived brightness, largely independently of changes in pitch or overall level [10]. Similarly, roughness—the perceptual attribute associated with beating or closely spaced partials—has been shown to contribute to judgments of musical tension and expressive intensity [11]. Recent overviews of timbre research describe the auditory system as organising sounds within a multidimensional space structured by such attributes [6], [12]. Taken together, these findings underscore a central principle of auditory perception: the human auditory system is a multidimensional analyser that parses sound into salient components rather than integrating all of its acoustic features equally.

Importantly for the present study, Ziemer's work and related psychoacoustic experiments show not only those perceptual dimensions exist, but that listeners can use those dimensions to perform tasks that have a procedural or decision-making character. Whether the task is distinguishing categories, matching auditory cues to data dimensions, or navigating toward a spatial target on the basis of perceptual differences, listeners demonstrate that they can extract and selectively attend to perceptual axes embedded

within sound [7], [8], [9]. This is not a passive reception of sound; it is a form of perceptually guided action.

From the perspective of audio-score theory, this psychoacoustic evidence provides a necessary foundation: if listeners can extract and act upon specific perceptual dimensions, then sound can function as a score to the extent that it carries structured, attended information along those dimensions. In the context of improvisation, this means that performers can treat particular aspects of sound as musical cues—to continue, to contrast, or to transform—based on the perceptual weight those dimensions carry for them. The next section examines how this capacity for selective perception underpins interaction in improvisational contexts and lays the groundwork for understanding how machine listening can emulate analogous processes.

Sound as a Score

The discussion above suggests that sound can function as a score only to the extent that it is selectively heard. In improvisation, this selectivity becomes a core compositional force. Accounts of ensemble practice repeatedly emphasise that listening and attentional focus are central organising principles of musical interaction: performers attend to salient cues in timbre, rhythm, and dynamics, and structure their responses accordingly [13]–[15]. Although they share the same sonic environment, musicians may attend to different aspects of what they hear, leading to different musical responses. This process is grounded not only in individual taste or style, but in the perceptual architecture of hearing itself, as described by auditory scene analysis and timbre research [3], [5].

From this perspective, an audio score is not a fixed sonic object that stands outside performance, but a pattern of selective attention that arises within it. A given sound contains many potential cues; it becomes a score when certain of those cues are treated as instructions. A sustained tone may function as a score for continuity, a noisy attack as a score for rupture, a particular spectral colour as a score for timbral imitation. What the score “is” depends on which dimensions of the sound are foregrounded in listening and how strongly they are weighted in guiding response. The score is thus embedded in the act of listening, rather than inscribed only in notation or recorded media.

This framing also opens a path toward AI-mediated improvisation. If audio scores operate through selective perception, then any system—human or algorithmic—that extracts and weights perceptual dimensions can, in principle, participate in audio scoring. Machine listening models, such as those used in concatenative synthesis, implement their own forms of selective hearing by analysing sound into descriptors and evaluating similarity in a feature space. When such a system is coupled to human improvisers, the sound they produce becomes a score for the machine, and the machine’s output, in turn, becomes a score for the performers. Audio scoring then takes the form of a continuous exchange of perceptual cues, in which human and algorithmic listeners guide one another over time.

In this sense, the audio score is no longer a one-way instruction from composer to performer, but a dynamic communicative process that unfolds through reciprocal listening. The composer’s role shifts toward designing the

conditions under which this reciprocal audio scoring can occur: choosing which aspects of sound will be analysable, how strongly they will be weighted, and how they will be made audible to both human and machine. The following chapter develops this idea by treating concatenative machine listening as a computational analogue of selective, feature-based listening, and by examining how such systems can be composed to act as partners in improvisation rather than as passive processors of sound.

3. MACHINE LISTENING AS AUDIO-SCORE INTERPRETATION

Machine Listening as Perceptual Filtering

The previous section argued that an audio score does not reside in a sound object alone, but in the *perceptual filtering* enacted by listeners: certain dimensions of sound are treated as significant, others recede, and musical action follows those perceptual weights. In this section, I extend that argument from human listeners to machine listeners, asking how similar logics of selection and salience operate within algorithmic systems.

Machine listening systems do not perceive sound holistically. They extract and weight descriptors corresponding to timbral, temporal, and energetic dimensions. In music information retrieval and computer audition, such descriptors are explicitly motivated by psychoacoustic and timbral research and are intended to approximate perceptually relevant aspects of sound [16], [17]. Corpus-based concatenative synthesis, for example, builds on databases of segmented and descriptor-analysed sounds, selecting units according to proximity to a target position in descriptor space. This process constitutes a form of perceptual filtering analogous to selective listening in human improvisation. Rather than treating the audio signal as a continuous physical waveform, machine-listening models transform sound into a set of features—such as MFCCs, spectral centroid, temporal envelope, and energy—that approximate how humans experience qualities like brightness, gesture, and intensity [17]–[20]. What the system “hears” is therefore not sound itself, but a structured representation of perceptual attributes.

The choice of descriptors determines which aspects of the sound are considered meaningful and which are ignored. A system that privileges MFCCs foregrounds timbral colour; one that emphasises envelope descriptors foregrounds gesture and articulation; a model centred on energy responds primarily to dynamics and salience. Work on descriptor sets for timbre analysis has shown that different combinations of features capture different facets of perceived timbre and are more or less effective depending on the task [18], [21]. Machine listening is thus already an interpretive act: it defines in advance what counts as musically or perceptually relevant within the sonic field. In this sense, the architecture of the listening model functions as a hidden score that shapes all subsequent musical decisions.

From the perspective of audio-score theory, this filtering process closely mirrors the way human performers attend selectively to aspects of sound in improvisation. Just as an improviser may follow brightness rather than pitch, or density rather than rhythm, a machine-listening system follows the dimensions encoded in its feature set. In CBCS such as CataRT, navigation through the corpus is driven by distances in descriptor space, making different descriptors and weightings directly audible as different “ways of listening” to the same material [18], [22]. The system’s “attention” is distributed according to mathematical weights rather than embodied intuition, yet the structural logic is comparable: both humans and machines respond not to the totality of sound, but to those dimensions that have been marked—by design or by habit—as salient.

Composing Listening: Machine Audition as Audio Score

If machine listening operates as a form of perceptual filtering, then the role of the composer in AI improvisation shifts from writing sounds to designing how sounds will be heard. In concatenative and corpus-based systems, the composer determines which descriptors are extracted, how these descriptors are weighted, and how similarity is evaluated. These decisions define the system’s mode of attention and therefore its musical behaviour. Composition becomes the construction of an auditory perspective rather than the specification of musical events [18].

This shift is clearly articulated in Pierre Alexandre Tremblay’s work with interactive concatenative systems such as *Sandbox*. In “Surfing the Waves: Live Audio Mosaicing of an Electric Bass Performance as a Corpus Browsing Interface,” Tremblay and Schwarz describe how an electric bass performance is analysed in real time and mapped onto a corpus of samples using CBCS, with “multi-dimensional mapping” and descriptor weighting central to expressive control [2]. The machine does not simply reproduce fragments from a database; it responds to the performer’s sound through a descriptor-based representation that maps multidimensional aspects of performance onto a corpus. I argue that this descriptor-based representation functions as a perceptual model that privileges certain timbral or gestural aspects. Because matching is determined by descriptor selection and weighting, different weighting schemes produce different relationships between live input and corpus response. Tremblay and Schwarz note that both the choice of descriptors and their relative weights are crucial to the perceived responsiveness and expressivity of the system. The composer, in this context, writes the listening strategy that governs the dialogue between human and machine.

Such practices reveal that in AI improvisation the audio score is not located in a fixed sonic object but in the configuration of machine audition. The performer’s sound functions as a score only because the system has been prepared to hear it in particular ways. Tremblay’s approach demonstrates that altering descriptor weights can transform the same input into radically different musical outcomes, much as two human improvisers might respond differently depending on whether they attend to rhythm, timbre, or dynamics [2]. The compositional act lies in

choosing which of these dimensions will carry prescriptive force.

Similar ideas appear in the work of other composers employing machine listening. Aaron Einbond treats real-time audio analysis and corpus-based matching as a means of articulating timbral relationships that could not be specified through notation alone, using descriptor spaces as “timbre spaces” for analysis and composition [23], [24]. Rodrigo Constanzo’s *C-C-Combine*, a corpus-based audio mosaicking instrument in Max/MSP, similarly behaves as a responsive partner whose actions are guided by feature-based similarity between live input and pre-analysed corpus grains; Constanzo explicitly highlights the expressive role of feature weighting in shaping the matching process. Across these practices, the machine is conceived not as an automated performer of prewritten material but as a listener whose perceptual orientation has been composed in advance.

Understanding machine listening in this way allows the concept of audio score to be extended beyond human performers. The input sound becomes a prescriptive signal that activates a designed mode of hearing within the system; the corpus provides the potential vocabulary; and the weighting scheme functions as the score that mediates between them. Composition, therefore, consists in shaping the conditions under which sound will influence sound: in short, the composer writes the ears of the machine. This perspective forms the basis for the case study that follows. In my own work with concatenative synthesis, different listening modes were constructed by altering descriptor sets and weights, enabling the same input to generate contrasting musical behaviours. The piece does not begin with a fixed sonic plan but with a designed auditory perspective through which the machine interprets the audio score presented by human performers.

4. CASE STUDY

OVERLAPPING (2025-26)—A CONCATENATIVE AUDIO SCORE FOR ALTO SAXPHONE AND ELECTRONICS

The preceding sections have proposed that in AI improvisation the audio score is constituted not by a fixed sonic object but by a designed mode of listening. Machine audition—through its descriptors, weights, and distance models—functions as an interpreter that decides which aspects of sound will acquire prescriptive force. The present case study examines how this concept is put into practice in my composition *Overlapping (2025-26)* for alto saxophone and live electronics, in which a machine-listening system built in Max/MSP interacts with the performer through multiple, parallel listening perspectives.

In this piece, audio scoring is realised as a reciprocal process: the saxophonist’s sound functions as a score for the machine, and the machine’s responses in turn function as a score for the performer. Rather than following a fixed timeline or notated structure, both agents rely on selective

listening. The system analyses the incoming sound into descriptors and generates output according to its current listening mode, while the performer shapes their playing in response to the evolving electronic texture. *Overlapping* thus serves as a practice-based exploration of how listening perception—human and algorithmic—can be organised so that sound itself becomes the medium through which human and AI improvisers guide one another over time.

Listening Modes and Empirical Observations

As mentioned before, the core assumption of concatenative systems is that similarity in perceptual feature space corresponds to meaningful musical continuity: audio fragments that are “close” in descriptor space produce coherent sequences, and those that are “far” create contrast or collage. In practice, however, this continuity depends not only on the incoming sound but also on the structure of the corpus against which it is matched. To explore this in relation to *Overlapping*, I conducted a series of informal tests using a common input—live alto saxophone—and two contrasting corpora: one built from field recordings, and one built from instrumental sounds.

Field-recording corpus

The first corpus consisted primarily of field recordings with complex, often diffuse spectral and textural characteristics. When saxophone material was used as input, the system tended to respond most convincingly to noise-based playing: breath noise, key clicks, inharmonic multiphonics, and other extended techniques with broad-band spectra. In these cases, the matching algorithm found neighbours in the field corpus that echoed the gestural envelope and rhythmic articulation of the saxophone while maintaining an environmental or textural character. The result was a concatenated output that, while not pitch-related, followed the shape of the input gesture and embedded it within an environmental texture.

By contrast, clearly pitched saxophone sounds—sustained tones or articulated melodic figures—matched less convincingly in this corpus. The strong periodic structure, clear harmonic content, and discrete pitch events of the saxophone did not find close analogues in the more diffuse, multi-source field recordings. The system still produced output, but the relationship between input and corpus felt weaker: the matches tended to ignore pitch and instead latch onto coarse spectral or energy similarities, which made the connection between performer and electronics less perceptible.

Instrumental corpus

The second corpus was built from instrumental sounds, including saxophone and other pitched sources, with relatively clear envelopes and harmonic or quasi-harmonic spectra. With this corpus, both noise-based and pitched saxophone materials produced coherent results, but in different ways. Noise gestures matched to noisy or spectrally dense instrumental sounds, again preserving envelope and rhythmic profile. Pitched gestures, however, now found

neighbours with comparable spectral centroid trajectories, partial structures, and onset patterns, resulting in concatenated sequences that reflected not only the input’s gesture but also its pitch-related structure.

In this instrumental corpus, the system’s responses were more clearly “musical” in the sense of perceptual coherence: listeners could more easily hear how the electronics followed or transformed the performer’s lines. Noise materials yielded textural and gestural continuity; pitched materials yielded timbral and quasi-melodic continuity.

Interpretation and implications

These observations suggest that concatenative listening is not neutral: it privileges those acoustic patterns that are both distinct in feature space and well represented in the corpus. When the corpus consists of field recordings, the system is particularly sensitive to broad-band, noisy, and textural input, and tends to disregard fine-grained pitch information. When the corpus is instrumental, the same input material can support both gestural and pitch-related continuity.

From an audio-score perspective, this behaviour aligns with the idea that a score operates through perceptual salience. The field-recording corpus lacks the kind of stable, discrete pitch events that improvisers often use as structural cues, so the machine listener treats the entire field primarily as a texture region. The instrumental corpus, by contrast, contains discrete, articulated sound objects that the system can cluster in feature space in ways that are more directly audible as correspondence with the saxophone’s playing.

At the same time, the situation is more complex than a simple field-versus-instrument dichotomy: segmentation, descriptor choice, and corpus diversity all influence the resulting matches. This paper does not aim to resolve these technical variables in detail. Rather, it takes these empirical observations as compositional information: they inform the selection and combination of corpora and listening modes used in *Overlapping*, so that the resulting machine listening supports an unfolding musical interaction rather than merely demonstrating a technique.

Sound Types and Performer Agency

Because *Overlapping* works with two contrasting corpora—one built from field recordings and one from instrumental sounds—the performer’s agency is articulated through a set of three sound types designed to probe how each corpus responds. These types are defined within a two-dimensional space whose x-axis represents temporal character (from static to gestural) and whose y-axis represents spectral character (from noise-based to pitched). Rather than treating the saxophone as a single source, the composition invites the performer to move within this coordinate space (Fig. 1), occupying regions associated with sustained static sounds, rhythmically articulated gestures, and fragmented static textures. Each region is perceived differently by the concatenative listener and therefore functions as a distinct audio score.

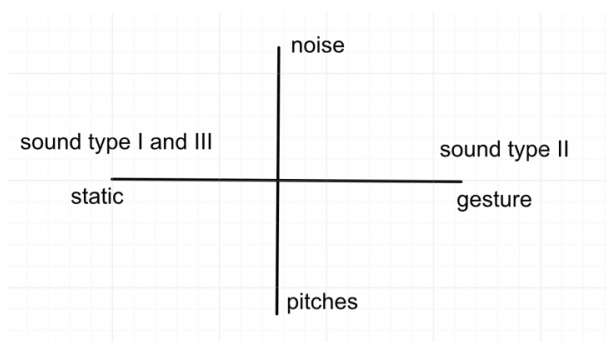


Figure.1. The coordinate for three types of input sound by alto Saxophone.

1. Static sustained sounds (Type I)

The first sound type consists of sustained, relatively static sounds: long pitched tones, airy breath sounds, or stable multiphonics. These sounds produce slowly changing descriptor values over time. When used as input, they tend to elicit outputs that preserve long envelopes and gradual spectral motion. With the instrumental corpus, sustained pitches encourage colour-matching responses, in which the system follows the spectral shape of the saxophone and produces quasi-harmonic continuations. With the field-recording corpus, sustained airy or noisy tones match more readily, and the system embeds the saxophone's sound within broader environmental textures. In both cases, the static temporal profile makes this type a score for continuity, while the choice between pitch and noise steers whether that continuity is timbral/melodic (instrumental corpus, pitched) or textural (field corpus, noisy/airy).

2. Gestural rhythmic sounds (Type II)

The second sound type is gestural and rhythmically articulated: tongued figures, slap attacks, key clicks, and rapid sequences that create a strong sense of pulse or kinetic energy (Fig.2). These gestures generate pronounced changes in onset, envelope, and short-term energy, making temporal descriptors highly salient. The concatenator responds to these variations with fragmented or event-like sequences that closely mirror the gesture's contour. With the instrumental corpus, pitched gestures often result in articulated, quasi-melodic mosaics that track both rhythm and rough pitch shape; with the field corpus, noisy or percussive gestures tend to produce bursts of environmental or textural material aligned to the same rhythmic pattern. Type II thus functions as a score for articulation and dynamism, inviting the system to track the morphology of the performer's gesture more than its sustained spectral colour.



Figure.2. Draft of sound type II.

3. Fragmented static textures (Type III)

The third sound type occupies an intermediate position: it is fragmented yet globally static. Here the performer uses short, vivid sound objects—isolated multiphonics, repeated key clicks, small pitch inflections—arranged in a way that does not build toward a clear rhythmic line but rather produces a flickering, pointillistic texture (Fig.3). Temporally, this type is less continuous than Type I but less directional than Type II; spectrally, it can be realised as either pitched or noisy material. For the concatenator, these inputs often produce outputs that echo the local micro-gestures while maintaining an overall sense of textural stasis. In the instrumental corpus, the system tends to pick out clusters of similar instrumental fragments; in the field corpus, it responds with small grains of environmental sound that punctuate the texture. Type III therefore shares with Type I a broadly static temporal quality, but its internal fragmentation makes it a score for textural detail rather than sustained continuity.



Fig.3. Draft of sound type III.

Noise and pitch as perceptual control

Across all three sound types, the performer also controls the spectral character of the input by moving between pitched and noise-based playing. Pitched sounds—conventional tones, microtonal inflections, stable multiphonics—provide periodic spectral organisation that is captured by descriptors such as spectral centroid and MFCC patterns, inviting responses that emphasise timbral resemblance and quasi-melodic continuity, especially in the instrumental corpus. Noise-based sounds—breath noise, inharmonic multiphonics, key clicks—foreground roughness and spectral diffusion, activating regions of the descriptor space that favour textural collage and contrast, particularly in the field-recording corpus.

By combining these three sound types with the pitch/noise distinction, the performer can intentionally steer the electronics. A sustained pitch (Type I) tends to generate elongated, colour-matching responses in the instrumental corpus; a gestural pitch (Type II) produces articulated, quasi-melodic fragments; a sustained noise (Type I realised noisily) encourages dense environmental

textures from the field corpus; and a fragmented noise texture (Type III) triggers pointillistic collage. The saxophonist therefore composes the audio score in real time by choosing not only *what* to play, but *how* it unfolds temporally and spectrally in relation to the two corpora.

Three Microphones as Three Ears

To open multiple pathways of interaction between performer and system, *Overlapping* employs three microphones, each routed to an independent concatenative process in Max/MSP. The three saxophone sound types described above—static sustained sounds, gestural rhythmic sounds, and fragmented static textures—are not tied to specific microphones in a fixed way; instead, the performer chooses in the moment which sound type to send to which input. In this sense, the microphones function less as technical channels and more as three potential ears for the machine listener. Figure 4 presents an excerpt of the sight-score for the improviser, outlining a possible unfolding of the three sound types in time and space. The vertical axis, read from bottom to top, functions as a timeline, while the three microphones are arranged from left to right to mirror their physical positions on stage. In this way, the diagram maps the performer’s movement between microphones onto both temporal progression and spatial orientation during the performance.

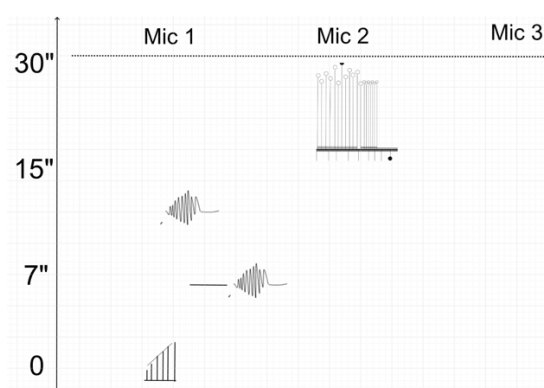


Figure 4. Sight-score for the idea of the improvisation.

Crucially, these ears are not defined by simple labels such as “loudness channel” or “spectral channel.” Although the underlying concatenative processes share a similar analytical structure, each microphone–corpus–descriptor combination tends to produce a characteristic response. Over the course of rehearsal and performance, these responses become recognisable to the performer: one microphone might habitually yield smoother, more continuous textures, another might produce fragmented or noisy collage, and a third might oscillate between the two depending on the input. The distinction between the three ears is therefore experiential rather than pre-defined. Each microphone becomes a different audio score because it triggers a different way of listening on the part of the machine.

The compositional design of *Overlapping* seeks to translate the theoretical idea of audio score into a concrete performance situation in which human and machine listeners influence one another. Rather than providing the saxophonist with a traditional notated score, I offer a sight-score of instructions (Figure 4) that describes how sound types are to be produced and how they may be routed to the three microphones. The written material does not prescribe musical results; it prescribes modes of sounding and routing that will function as scores for the machine listeners.

In performance, the improviser chooses among these microphones according to the evolving dialogue with the electronics. A sustained tone might be sent to the ear that tends to produce continuity, while a noisy gesture might be directed to the ear that usually provokes rupture. Through this exploratory process the musician learns how each ear tends to “answer” and adapts their playing accordingly. The machine, in turn, responds to the performer’s choices with new sonic material that feeds back into the performer’s listening and decision-making. Audio scoring thus becomes a looped process: the saxophone’s sound scores the concatenators, the concatenators’ output scores the performer, and both adjust their behaviour in response.

This design emphasises that in AI improvisation the score is not fully determined in advance. The microphones provide entry points into the system, but their musical meaning arises only through interaction. Feedback from the concatenators guides the performer toward new sound choices, and these choices in turn redefine what the machine hears. The audio score is therefore an emergent relation among performer, microphones, and machine listening rather than a fixed mapping of descriptors.

5. CONCLUSIONS

This paper has proposed a transformation of the concept of audio score from an instruction toward pure sound filtering¹. Building on practices articulated by Louise d’Heudière, where sound functions as an instruction for performers, I have argued that in AI-mediated improvisation sound can become a score not by representing music, but by actively shaping how music is heard and generated. The audio score, in this sense, is neither a notation nor a fixed object, but a dynamic relationship between sound, listener, and system.

Psychoacoustic research demonstrates that listeners do not perceive sound as an undifferentiated whole but extract salient dimensions such as brightness, roughness, tonalness, and gesture. Improvisation operates through this selective attention: performers choose which aspects of the sonic environment will guide their responses. I have suggested that concatenative machine listening embodies an analogous process. By analysing sound into perceptual descriptors and weighting these descriptors within a feature space, the system performs a form of geometric listening in which similarity, continuity, and collage emerge from computational judgments of perceptual proximity.

¹ Here the pure sound indicates the sound of non-instruction.

The case study of *Overlapping* for alto saxophone and electronics illustrates how this idea can be realised compositionally. Through the use of multiple corpora, three sound types, and three microphones routed to separate concatenative processes, the piece constructs several parallel “ears” that interpret the performer’s sound in distinct ways. The written instruction functions only as a meta-score; the true score is the sounding interaction between saxophonist and concatenators. In performance, both the acoustic instrument and the electronic responses are audible to the audience, making the audio score perceptible as an evolving dialogue.

At the same time, the case study reveals important limitations of concatenative machine listening as an audio-score interpreter. While the system can respond convincingly to particular aspects of the saxophone sound—such as envelope shape, spectral colour, or energy—it cannot apprehend the sound object as a whole. Each concatenator listens through a restricted set of descriptors, and what lies outside those descriptors remains inaudible to the machine. In practice, no single analytical pathway was sufficient to represent the richness of the instrument; the three-microphone setup emerged from the necessity to distribute listening across multiple, partial perspectives. The audio score interpreted by the system is therefore not the sound itself, but a projection of that sound onto chosen analytical frames.

Recognising this constraint is part of composing with machine listening, but it also becomes a creative resource. Because each object hears differently, the performer can navigate among these partial perspectives, treating them as alternative scores. The dialogue between human and machine thus exposes the constructed nature of listening: neither the saxophonist nor the concatenator possesses a complete understanding of the sound, yet their interaction generates musical form. Composition becomes the design of listening conditions rather than the prescription of musical events. To compose with concatenative systems is to compose the ears of the machine—to decide which dimensions of sound will acquire authority and how these dimensions will influence further sound.

Future work might explore more flexible listening models in which descriptors are dynamically reweighted or combined, allowing the machine to shift its attention as human improvisers do. Hybrid systems that integrate learning-based representations with concatenative methods may also offer richer interpretations of audio scores, and ensembles could engage with multiple machine listeners simultaneously, each offering a different reading of the same performance. More broadly, understanding audio scores as forms of perceptual agency invites a reconsideration of what it means to write music in an age where machines listen. In this context, sound becomes not simply the end of composition, but its beginning.

6. REFERENCES

- [1] L. P. M. L. d’Heudières, “Scores in Time: Exploring the Use of Sound as a Medium for Notation in Musical Composition,” Ph.D. dissertation, Bath Spa Univ., Bath, U.K., 2020.
- [2] P. A. Tremblay and D. Schwarz, “Surfing the Waves: Live Audio Mosaicing of an Electric Bass Performance as a Corpus Browsing Interface,” in Proc. Int. Conf. New Interfaces for Musical Expression (NIME), Sydney, Australia, 2010, pp. 447–450.
- [3] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA, USA: MIT Press, 1994.
- [4] J. M. Grey, “Multidimensional perceptual scaling of musical timbres,” *J. Acoust. Soc. Am.*, vol. 61, no. 5, pp. 1270–1277, May 1977, doi: 10.1121/1.381428.
- [5] S. McAdams, “The perceptual representation of timbre,” in *Timbre: Acoustics, Perception, and Cognition*, K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, and R. R. Fay, Eds., *Springer Handbook of Auditory Research*, vol. 69. Cham, Switzerland: Springer, 2019, pp. 23–57, doi: 10.1007/978-3-030-14832-4_2.
- [6] K. Siedenburg and S. McAdams, “Four distinctions for the auditory ‘wastebasket’ of timbre,” *Frontiers in Psychology*, vol. 8, Art. no. 1747, 2017, doi: 10.3389/fpsyg.2017.01747.
- [7] T. Ziemer, “A psychoacoustic auditory display for navigation,” in Proc. Int. Conf. Auditory Display (ICAD), 2018.
- [8] T. Ziemer and H. Schultheis, “Three orthogonal dimensions for psychoacoustic sonification,” *arXiv preprint arXiv:1912.00766*, 2019.
- [9] T. Ziemer, D. Black, and H. Schultheis, “Psychoacoustic sonification design for navigation in surgical interventions,” in Proc. Meetings on Acoustics, vol. 30, no. 1. Melville, NY, USA: Acoustical Society of America, 2017.
- [10] E. Schubert and J. Wolfe, “Does timbral brightness scale with frequency and spectral centroid?” *Acta Acustica united with Acustica*, vol. 92, no. 5, pp. 820–825, 2006.
- [11] M. M. Farbood, “A parametric, temporal model of musical tension,” *Music Perception*, vol. 29, no. 4, pp. 387–428, 2012.
- [12] K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, and R. R. Fay, Eds., *Timbre: Acoustics, Perception, and Cognition*. Cham, Switzerland: Springer International Publishing, 2019.

- [13] P. Oliveros, *Deep Listening: A Composer's Sound Practice*. New York, NY, USA: iUniverse, 2005.
- [14] G. E. Lewis, "Interacting with latter-day musical automata," *Contemporary Music Review*, vol. 18, no. 3, pp. 99–112, 1999.
- [15] I. Monson, *Saying Something: Jazz Improvisation and Interaction*. Chicago, IL, USA: University of Chicago Press, 2009.
- [16] X. Serra, "Musical sound modeling with sinusoids plus noise," in *Musical Signal Processing*, C. Roads, S. Pope, A. Piccilli, and G. De Poli, Eds. Lisse, The Netherlands: Swets & Zeitlinger, 1997, pp. 91–122.
- [17] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The Timbre Toolbox: Extracting audio descriptors from musical signals," *J. Acoust. Soc. Am.*, vol. 130, no. 5, pp. 2902–2916, 2011.
- [18] D. Schwarz, "Corpus-based concatenative synthesis," *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 92–104, 2007.
- [19] D. Schwarz, "Concatenative sound synthesis: The early years," *J. New Music Res.*, vol. 35, no. 1, pp. 3–22, 2006.
- [20] B. Logan, "Mel frequency cepstral coefficients for music modeling," in *Proc. Int. Symp. Music Inf. Retrieval (ISMIR)*, 2000.
- [21] I. Czedik-Eysenberg and C. Reuter, "Evaluating timbre-related audio descriptors across different libraries and multimodal embeddings," in *Proc. DAS | DAGA 2025: 51st Annual Meeting on Acoustics*, 2025, pp. 171–174.
- [22] A. Einbond, C. Trapani, A. Agostini, D. Ghisi, and D. Schwarz, "Fine-tuned control of concatenative synthesis with CATART using the BACH library for MAX," in *Proc. Int. Comput. Music Conf. (ICMC)*, Athens, Greece, Sep. 2014.
- [23] A. Einbond, T. Carpentier, D. Schwarz, and J. Bresson, "Embodying spatial sound synthesis with AI in two compositions for instruments and 3D electronics," *Comput. Music J.*, vol. 46, no. 4, pp. 43–61, 2022, doi: 10.1162/comj_a_00664.
- [24] A. Einbond, J. Bresson, D. Schwarz, and T. Carpentier, "Instrumental radiation patterns as models for corpus-based spatial sound synthesis: *Cosmologies* for piano and 3D electronics," in *Proc. Int. Comput. Music Conf. (ICMC)*, San Francisco, CA, USA, 2021, pp. 148–153.

GENERATIVE AUDIOSCORES AS ENGINES OF SOCIAL CHOREOGRAPHY IN „VILLANELLES DE VOYELLES II“

Sandeep Bhagwati

Concordia University Montréal
sandeep.bhagwati@concordia.ca

Kasey Pocius

CIRMMT, Concordia University Montréal
kaseypocius@concordia.ca

ABSTRACT

This paper analyzes an instance of a generative audioscore through an analysis of “*Villanelles de Voyelles II*” (2022), a non-linear audio-based score system. Building on earlier work on Elaborate Audio Scores, the paper argues that this piece represents a qualitative shift from audio scores as conveyance mechanisms toward procedurally composed dramaturgical systems. Musical form emerges from a constrained generative architecture combining weighted instruction sets, elastic temporal distributions, relational pitch logic, and periodic system-wide interventions. Beyond sonic and temporal organisation, the paper foregrounds the affordances of generative audioscores for spatial movement and emergent theatricality. By encoding relations of attention, proximity, and response rather than choreographic gestures, the audioscore enables performers to enact socially legible behaviours without representational intent. The paper situates generative audioscores as notational systems that compose sound, time, space, and social interaction.

1. INTRODUCTION: FROM ELABORATE TO GENERATIVE AUDIO SCORES

In 2017, Sandeep Bhagwati had introduced the concept of the **Elaborate Audio Score** (EAS) as a type of situational score that conveys compositional intent primarily through sound, rather than through visual symbols, graphical representation or tactile sensation. [1] These scores often employ headphones as their interface and rely on real-time auditory messages—such as instructions, cues, sound examples, and relational prompts—to coordinate ensemble performance. A central claim of that paper was that audio scores afford modes of musical conveyance, bodily freedom, and temporal elasticity that are difficult or impossible to realise with visual notation alone.

However, that earlier framework left one central compositional question largely open: how can large-scale form, dramaturgical coherence, and perceptible intentionality be composed in an audio score without reverting to linear scripting, fixed sequences, or conductorial intervention?¹ In

¹ Non-linear and open notational strategies have a long history in

other words, how can an audio score remain entirely situative and open while still exhibiting the hallmarks of composed form?

This paper proposes one conception of a generative audioscore as a response to this question. Through an analysis of *Villanelles de Voyelles II* (2022), we argue that generative audioscores constitute a distinct notational paradigm in which compositional intent is embedded not in event sequences, but in probabilistic structures, temporal envelopes, and relational constraints. These structures enable performances that are repeatable at the level of dramaturgical behaviour, yet irreducible at the level of sonic and spatial detail.²

Beyond sound and time, this paper advances a further claim: generative audioscores can function as engines of procedural social choreography, enabling performers to enact socially and theatrically legible behaviours—such as conversational clustering, canvassing, or collective deliberation—without representational intent or explicit theatrical instruction. This introduces a layer of non-representational theatrical affordance that emerges from the interaction between audio instruction, temporal elasticity, and free spatial movement.

2. CONTEXT: VILLANELLES DE VOYELLES (VERSION 2)

Villanelles de Voyelles is part of a longer cycle of works by Bhagwati that engage with the Indian tradition of nir-gun bhajans: spiritual songs addressed to nameless, non-anthropomorphic entities. In this piece, singers produce

twentieth-century Eurological composition, particularly in visual and spatial score formats. Graphic and multi-dimensional scores by composers such as John Cage, Pierre Boulez, Karlheinz Stockhausen, Iani Christou, and later John Zorn, allowed for non-sequential reading paths, performer choice, and indeterminate ordering through two- or three-dimensional visual layouts. In contrast, the audioscore unfolds in real time along a single temporal axis and cannot be surveyed, re-ordered, or navigated spatially by the performer. This structural linearity poses a distinct challenge for composing non-linearity, requiring the use of generative processes, probabilistic branching, and relational dependencies to achieve degrees of openness that visual scores can realise through spatial organisation alone.

² A first work by Bhagwati (software: Joseph A. Browne, Travis West; hardware: Alex Bachmayer) exploring such dramaturgical uses of situative scores was “Silence Not Absence” (first performance at TENOR 2018) for stationary cellist (Elinor Frey) and moving trombonist (Felix Del Tredici). Viewers of this and later performances frequently compared this piece to a theatrical/choreographic relationship drama between the two musicians, arising from the interplay between playing and moving instructions. While this work did use audio scores to provide musicians with primary musical material, the movement instructions were transmitted separately via a tactile score interface (body:suit:score), which was controlled live by the audience via a smartphone app. Movement and sound were thus not structurally or agentially connected. For this reason, this work is not further discussed here.

vowel-based vocal sounds, avoid consonants almost entirely, and occasionally mimic bird calls or environmental sounds. The performers are ideally masked and presented as non-human or more-than-human entities, situated between ritual, song, and abstraction.

The first version of *Villanelles de Voyelles* (2017) had employed a fixed, linear audioscore composed on a digital audio workstation. While already freeing performers from visual notation and enabling bodily movement, this version still relied on a predetermined sequence of instructions. In 2021–22, the piece was reconceived as a non-linear, generative audioscore, developed as a flowchart instruction booklet by Bhagwati and realised as software app in MAX 8 by Kasey Pocius. The first performances were realized by the Neue Vocalsolisten, first on May 10 2022, at TENOR 2022 in Marseille and soon after at the Theaterhaus Stuttgart (June 1). Version II thus replaces linear sequencing with a rule-based system that generates instructions in real time for each singer, coordinated through probabilistic weighting, temporal distributions, and occasional system-wide constraints. Each performance unfolds differently, yet exhibits recognisable formal and behavioural characteristics across iterations (more on that below).

3. FROM INSTRUCTIONS TO SYSTEMS

At the surface level, the audioscore of *Villanelles de Voyelles* consists of a clearly delimited instruction vocabulary. Vocal behaviours are defined through combinations of voice techniques and dynamic or pitch envelopes, supplemented by ensemble-related instructions governing imitation, accompaniment, attention, motion, and spatial orientation.

In isolation, such an instruction set could function as an improvisational toolkit. What transforms it into a improvisation is not the content of individual instructions, but the system that governs their appearance, timing, and coordination. In Version II, the audioscore ceases to be a sequence of messages and instead becomes a procedural environment in which instructions emerge according to compositional logic. This marks a conceptual shift: the score no longer tells performers what happens next, but defines how things may happen, how often, in which relations, and with what degree of coordination. The notated object is no longer an event stream, but a space of constrained possibilities.

This shift is crucial for understanding the theatrical and social affordances discussed later in the paper. Because performers cannot anticipate the next instruction visually, and because instructions arrive asynchronously and elastically, performers must remain in a state of heightened situational awareness. Their responses necessarily integrate sound, movement, gaze, and social orientation.

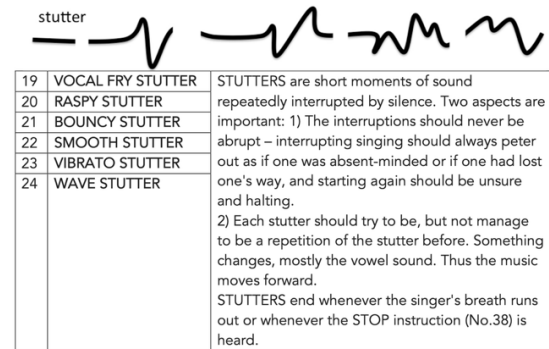
4. THE “VILLANELLES DE VOYELLES” GENERATIVE AUDIOSCORE

The audioscore of *Villanelles de Voyelles II* operates as a real-time generative system that produces up to 8 performer (or performer grouping) -specific instruction streams

under a shared set of probabilistic and dramaturgical constraints. Rather than generating musical events directly, the system generates instructions whose sonic, spatial, and social realisation is left to the performers. This section outlines the technical logic by which instructions are selected, timed, coordinated, and occasionally overridden.

4.1 Instruction Vocabulary and Sub-Bag Structure

The complete instruction set consists of 55 discrete instructions, divided into two functional groups.



19	VOCAL FRY STUTTER	STUTTERS are short moments of sound repeatedly interrupted by silence. Two aspects are important: 1) The interruptions should never be abrupt – interrupting singing should always peter out as if one was absent-minded or if one had lost one's way, and starting again should be unsure and halting. 2) Each stutter should try to be, but not manage to be a repetition of the stutter before. Something changes, mostly the vowel sound. Thus the music moves forward. STUTTERS end whenever the singer's breath runs out or whenever the STOP instruction (No.38) is heard.
20	RASPY STUTTER	
21	BOUNCY STUTTER	
22	SMOOTH STUTTER	
23	VIBRATO STUTTER	
24	WAVE STUTTER	

Figure 1. Example of a vocal instruction legend.

First, **36 vocal behaviour instructions** are generated through the combinatorial pairing of six basic voice timbre techniques (e.g. vocal fry, raspy, smooth) with six dynamic or pitch envelopes (drone, bell, hits, stutter, slump, rise). These instructions define continuous vocal behaviours rather than discrete events and always remain active until explicitly terminated.

Second, 19 ensemble and behavioural instructions regulate stopping, mimicking, accompanying, motion, turning, and speech-like actions. These instructions often redirect attention toward other performers or toward environmental sound, and several explicitly invite spatial movement.

For generative purposes, all instructions are grouped into thematic sub-bags (e.g. drones, bells, hits, motion, turning, singing with pitch-class logic). Each sub-bag is assigned a probabilistic weight that determines how frequently instructions from that category are selected over time

4.2 Multi-Stage Instruction Sampling

Each instruction issued to a performer results from a multi-stage sampling process rather than a single random choice. This process unfolds independently for each voice:

1. *Sub-bag selection*: The system samples one sub-bag from the full instruction bank, using weighted probability distributions. This step shapes the long-term behavioural balance of the piece (e.g. prevalence of reactive versus sustained behaviours).
2. *Instruction selection within sub-bag*: Within the chosen sub-bag, a specific instruction is selected by plain random sampling without replacement, preventing immediate repetition and promoting local variation.

Sandeep Bhagwati VILLANELLES DE VOYELLES – generative improvisation score - inner logic sheet 1

Each complete instruction will require 3 or 4 random sampling procedures.

SAMPLE 1 The entire bank of “Instructions” is a replaceable bag to sample one of the weighted sub-bags (see below).

SAMPLE 2 Within each SUB-BAG, the sampling is plain random and without replacement (*chooses instruction soundfile*)

SAMPLE 3 Within each INTERVAL, the sampling is Gaussian around median value (*chooses number of seconds until next choice*)

SAMPLE 4 [Only if 41/43] sample with replacement from the “Birds” Bag & [only if 54-55] sample following the “chords logic” from the “Pitch-Classes” Bag.

INSTRUCTIONS

	1-6	7-12	13-18	19-24	25-30	31-36	37-40	41-44	45-48	49-51	52-53	54-55	
SUB-BAG	Drone	Bell	Hits	Stutter	Slump	Rise	Stop/ Fade	React + 41/43-Birds	Motion	Turn	Asides	Sing + Pitch-Class	SUB-BAG
WEIGHT	9%	8%	7%	3%	5%	4%	9%	20%	6%	6%	6%	17%	WEIGHT
INTERVAL (sec)	21 ± 8	14 ± 5	10 ± 3	13 ± 4	6 ± 3	6 ± 3	15 ± 5	25 ± 10	7 ± 3	9 ± 4	11 ± 3	8 ± 3	INTERVAL (sec)

BIRDS

(only following instructions 43 or 45)

Any Soundfile 01-16

Chosen randomly
with replacement

PITCH-CLASS

(only following instruction 55)

- The first pitch-class chosen randomly in any track (p1) determines those chosen in the following tracks until all tracks have chosen once. Then the draw count begins anew (new iteration).
- Pitches are assigned to voices in the order of when they enter (or, with simultaneous chords: whose instruction would be up next).
- The +n or -n refers to the interval distance from the given pitch (1=min second up to 7=fifth etc).
- modulo 12: all sounds in the same octave, i.e. pitch-class, not real pitch. The singers can then choose any octave.
- P_n refers to P4 from the previous iteration (drawn randomly for the very first iteration).
- If there is a P8 (i.e. a version with 8 singers) P8 is chosen entirely at random.

Weight	P1	P2	P3	P4	P5 (optional)	P6 (optional)	P7 (optional)
50%	= P ₀ ± 3	= P1 ± 4	= P2 ± 3	= P3 ± 5	= P2	= P3	= P1
35%	= P ₀ ± 1	= P1 ± 2	= P1	= P3 ± 1	= P4 ± 2	= P5 ± 1	= P6 ± 3
15%	= P ₀	= P1 ± 7	= P2 ± 1	= P2	= P4 ± 6	= P4	= P5

Figure 2. Page 1 of the inner logic booklet of the generative score devised by Bhagwati for Pocius. Note how the Birds and Pitch-Class samples only come into play when specific instructions are chosen.

3. *Temporal interval sampling:* Each sub-bag is associated with a characteristic temporal interval. The time until the next instruction is sampled from a Gaussian distribution centred on a median value, introducing temporal elasticity while maintaining statistical regularity.
4. *Conditional secondary sampling:* Certain instructions trigger additional sampling procedures. For example, mimic and accompany instructions may draw from an auxiliary “Birds” soundfile bag, while pitch-based singing instructions invoke a separate pitch-class logic.

This layered sampling strategy allows the system to remain statistically stable at the macro level while producing unpredictable micro-level outcomes.

4.3 Pitch-Class Logic for Harmonic Coordination

Harmonic coordination is handled through a relational pitch-class system rather than fixed pitch assignment. When a singing instruction requiring pitch is generated, the first pitch-class in a cycle is chosen randomly. Subsequent pitch-classes for other voices are derived relationally from that initial choice, using weighted intervallic offsets (e.g. ±1, ±3, ±5 semitones modulo 12).

Pitch-classes are assigned to voices in the order of instruction entry, not by fixed voice identity. Once all voices have received a pitch-class, the system resets and begins a new pitch-class iteration. Singers are free to realise pitch-classes in any octave.

This approach ensures harmonic coherence across voices while avoiding both tonal progression and fixed chord structures, resulting in slowly shifting harmonic fields that re-

main perceptibly related across performers. Thus, pitch-class representation functions here as a relational listening guideline rather than articulating any historical tonal or stylistic commitment.

4.4 Dramaturgical Constraints and Forced Coordination

In addition to stochastic generation, the system periodically introduces global dramaturgical constraints that override local instruction streams. These forced events are applied simultaneously—or within a defined temporal spread—to all voices.

Forced constraints are themselves probabilistically scheduled and may include:

- coordinated singing (often using pitch-class logic),
- collective stopping or fading,
- shared turning or motion behaviours.

A spread parameter determines how tightly aligned these instructions are in time: a spread of zero produces simultaneity, while larger spreads distribute the same instruction across voices within a defined temporal window. Any locally generated instruction scheduled during this window is suppressed or displaced.

Following a forced event, the system applies a time shift to each performer’s instruction stream, displacing subsequent events forward in time. This prevents forced coordination from accumulating into metric regularity and preserves long-term temporal fluidity.

4.5 Temporal Continuity and Instruction Switching

A key design principle of the score logic is continuity of action. Performers are instructed never to stop vocalising

while listening to a new instruction; instead, they transition seamlessly from one behaviour to the next as soon as the audio message ends. This design choice ensures that instruction switching produces perceptible transformation rather than silence or rupture, and it supports the emergence of overlapping behaviours across the ensemble.

4.6 Performer Interface and System Control

The generative logic is implemented in a software interface that allows a “software conductor” to set global parameters such as number of performers, total duration, and ending mode (fade or stop). Once initiated, the system runs autonomously, with no further human intervention. Control is exercised entirely through probabilistic design rather than real-time decision-making. The “software conductor” thus has a supervisory but, except for troubleshooting, not a live-interactive role.

To enable the ‘temporal elasticity’ described earlier, we designed the system to initially presents the user with a setup window, which allows for the first initial compositional constraints to be placed. The facilitator inputs the overall duration of the piece, the number of parts to be generated (between four and eight), and the desired ending mode—either an explicit stopping instruction or a gradual fade-out. Additional technical options include a routing matrix that allows instruction streams to be assigned to specific output channels as required by the performance situation, as well as monitoring presets for rehearsal and preparation. Once the performance is initiated, these parameters are fixed and cannot be altered.

A solo mode in the interface, in addition to its monitoring function during performance, allows a performer to run a single instruction stream independently of the ensemble. This allows singers in preparation to familiarise themselves with the range of instruction types, their typical pacing, and the kinds of temporal situations they may encounter during performance, without rehearsing fixed sequences or outcomes. The solo mode thus supports embodied learning of the system’s behaviour rather than memorisation of material.

4.7 During Performance

The technical setup assumes multi-channel audio output, with one discrete channel per performer. Each channel is transmitted wirelessly to headphones worn by the singer. Crucially, the work is conceived for non-amplified performance: the singers’ voices are heard acoustically in the space, without reinforcement through loudspeakers. Any form of electroacoustic amplification—even spatialised systems such as wave field synthesis—would compromise the singers as well as the audience’s ability to localise sound sources accurately and would thereby disrupt the perception of spatial relations among and across performers. Because the work’s structural, social and theatrical affordances depend on everyone’s ability to attribute sounds unambiguously to moving bodies in space, the preservation of natural sound-source localisation is a structural requirement rather than a technical preference.

After activation, all parts are generated concurrently. This simultaneity is essential, as several stochastic processes—most notably those governing pitch selection for sustained singing and heartbeat-tempo singing—operate on relational dependencies across parts. Pitch-classes are not chosen independently but are weighted in relation to pitch-classes currently active in other voices and to those that have appeared earlier in the performance. Rather than producing parallel, autonomous streams, the system thus maintains a dynamically interdependent harmonic field.

Because pitch selection is relational rather than absolute, performers are encouraged to cultivate heightened situational awareness. They must attend not only to their own instruction stream but also to the sounding actions of other performers, whose pitches implicitly shape the harmonic affordances of subsequent instructions. Listening therefore functions as a structural necessity rather than an aesthetic preference.

Once an instruction has been delivered to a performer, the system stochastically samples a waiting interval before generating the next instruction for that part. These inter-instruction intervals are drawn independently for each performer, resulting in asynchronous instruction streams that nonetheless remain statistically coordinated through shared probability distributions and constraint logic.

A global master clock governs the overall temporal frame of the performance. It issues instructions marking the beginning and end of the piece and may introduce forced termination events—such as global stop or fade commands—that interrupt ongoing instructions or waiting intervals. Outside of these global interventions, a local logic layer ensures that instructions addressed to a single performer do not overlap, preserving perceptual clarity and continuity of action.

During performance, the overview interface displays, in real time, which instruction soundfiles are currently active in each performer’s headphones, as well as the most recently assigned pitch-class for each part, represented numerically from 1 to 12. This display serves for the diagnostic supervision of software functionality. It offers no means of real-time intervention, does not influence instruction generation, and does not constitute a conductorial control layer. Compositional agency resides entirely in the probabilistic structures and relational dependencies encoded in the system prior to performance.

Finally, the use of a generative audioscore raises the question of repeatability and recognisability. While each realisation of *Villanelles de Voyelles II* unfolds differently at the level of specific events, spatial trajectories, and sonic detail, audience experience tends to exhibit a high degree of recognisability across performances. This recognisability does not derive from the repetition of musical material, but from the recurrence of characteristic densities, behaviours, temporal flows, and social configurations.

One way to describe this is by analogy to everyday navigation through familiar urban spaces: although a street presents itself differently each time one walks along it—through changing light, weather, encounters, and incidental events—the overall experience of that street remains distinct from that



Figure 3. Overview window of the score. Note the dropdown menus which offer choices. 3:00 min is set as the minimum performance time. In seated audience settings, the piece works well at 8-12 minute duration, but it could also be realised e.g. as a day-long installation piece in a gallery. The Start/Stop buttons are useful during rehearsals or in emergencies, not in performance.



Figure 4. The supervision display of all audio channels: which instruction files are currently played and which pitches were last assigned to which channel (A=1, Bb=2 etc)

of other streets or other cities. Similarly, performances of *Villanelles de Voyelles II* are irreducible in their particulars, yet recognisable in their experiential profile. The work thus maintains identity not through fixed form, but through the consistent generation of a particular field of relations.

5. FORCED EVENTS AND COMPOSITIONAL AGENCY

One of the defining features of the generative audioscore in *Villanelles de Voyelles* is the alternation between stochastic flow and periodic, system-wide intervention. At irregular but compositely determined intervals, the system forces coordinated actions across all performers: shared stops, collective turns, harmonically related vocal events, or synchronised behavioural shifts.

These forced events function neither as cues from a conductor nor as narrative climaxes. Rather, they articulate moments of compositional agency within an otherwise self-organising field. The audioscore temporarily asserts a global will, only to withdraw again into probabilistic autonomy.

From a theatrical perspective, these moments often read as sudden collective attention shifts: a group freezing, re-orienting, or converging before dispersing again. Importantly, no symbolic meaning is attached to these actions; their perceptibility lies in their coincidence and shared timing, not in representational content.

This oscillation between autonomy and intervention is central to the piece's dramaturgy. It enables the emergence of perceptible form without prescribing narrative, character, or dramatic arc.

6. SPATIAL AGENCY AND EMERGENT THEATRICALITY

6.1 Audioscore as Enabler of Free Spatial Action

By replacing visual notation with individual audio instruction streams delivered via headphones, the audioscore liberates performers from fixed orientation, frontality, and visual coordination.[1] Performers are not required to maintain eye contact with a conductor, a score, or one another; instead, they navigate the performance space through listening, movement, and situational awareness. Spatial mobility is thus not an optional theatrical addition but a structural consequence of the score's mode of address. In *Villanelles de Voyelles*, spatial action is frequently prompted explicitly through motion-related instructions, but even in their absence, movement emerges as a practical response to listening, orientation, and acoustic interaction. The audioscore therefore establishes the conditions for spatial agency without prescribing spatial patterns.

6.2 Instruction Types and Emergent Social Kinetics

Several instruction families in the audioscore encode relations of attention and response rather than specific actions. Instructions such as accompany what you hear, mimic outside sounds, or move towards singer X define directional or relational tasks without determining their physical realisation. Performers must continuously translate these relational prompts into embodied behaviour using their own social and spatial judgement.

As a result, individual actions aggregate into recognisable but unstable configurations: temporary dyads, clusters, focal figures, or dispersed parallel activity. These configurations are not rehearsed, named, or assigned in advance.

They emerge from asynchronous instruction timing, performer movement, and reciprocal listening, forming what can be described as a field of “social kinetics” - understood here not as a theoretical concept but rather as the evolving patterns of movement, proximity, and orientation arising from relational instructions.

6.3 Social Choreography without Representation

The emergence of socially legible configurations without representational intent places *Villanelles de Voyelles* in dialogue with expanded notions of choreography developed in performance and dance theory. André Lepecki has argued that choreography operates beyond codified dance technique as a structuring of bodies, attention, and agency in space and time. [2, 3, Introduction, Chapter 1 & 2] Andrew Hewitt’s concept of *social choreography* further extends this perspective to everyday movement, describing how social relations become perceptible through patterns of gathering, dispersal, and proximity.[4]

A closely related theatrical precedent can be found in Peter Handke’s *Die Stunde da wir nichts voneinander wussten*, a play composed entirely of stage directions, which Handke likened to observing passers-by from a café on a public square.[5] Meaning arises not through narrative or character, but through coincidence, interruption, and the observers’ interpretation of fleeting social constellations. Samuel Beckett’s *Act Without Words I & II* similarly demonstrate how theatrical meaning can emerge from timed action and constraint without dialogue, psychology, or representation. [6]

In *Villanelles de Voyelles*, however, such configurations are neither scripted nor fixed. They arise through a real-time, rule-based system that continuously redistributes attention and action across multiple performers. To account for this mode of organisation, we propose the term **procedural social choreography**: a choreography not of movements or roles, but of conditions under which social relations become perceptible through listening, proximity, and timing.

6.4 Non-Representational Theatrical Affordance

Theatrical meaning in *Villanelles de Voyelles* does not stem from acted gesture, character, or narrative intention, but from the audience’s perception of evolving relations among performers. This aligns with non-representational approaches in performance theory, which emphasise meaning as an emergent effect of practice, perception, and co-presence rather than symbolic encoding.

Susan Leigh Foster has described choreography as a form of cultural scripting that renders bodily behaviour intelligible without requiring explicit representation.[7, Introduction & Ch.1] William Forsythe’s *Improvisation Technologies* offers a complementary choreographic model in which movement is generated through procedural prompts — spatial vectors, alignments, and attentional tasks—rather than fixed sequences. [8] While Forsythe’s work primarily addresses the organisation of individual movement, it demonstrates how procedural systems can produce coherent, recognisable behaviour without prescribing form.

In *Villanelles de Voyelles*, this logic is extended from individual movement to social interaction. The audioscore affords theatrical meaning by organising who listens to whom, when attention shifts, and how proximity changes over time. The resulting configurations are legible to audiences as social situations, yet they are not representations of such situations. Theatricality thus emerges as a **non-representational theatrical affordance** of the procedural system itself. Fig. 5 visualises the generative audioscore of *Villanelles de Voyelles* as a multi-layered process in which listening, rather than gesture or movement, functions as the primary motor of musical, spatial, and social action. Although organised as a left-to-right flowchart, the diagram includes feedback loops (red arrows) that indicate a recursive rather than linear process

7. FROM AUDIOSCORE TO EMERGENT SOCIAL MEANING

In fig. 5, the leftmost cluster represents the **generative audioscore system**. Instead of a sequence of events, it comprises an instruction vocabulary, a catalog of sound-files and pitch examples, probabilistic weighting, temporal distributions, and periodic system-wide constraints. These elements define a space of constrained possibilities from which instructions are generated in real time and delivered as private, one-to-one audio messages to each performer.

This instruction stream leads into the second cluster, **Listening to... audioscore** where performers parse not only semantic or sonic content but also timing, urgency, and relational cues. Because instructions and sonic cues arrive in real time and cannot be previewed, listening already entails heightened situational attention. From here, attention shifts to **Listening to... Other Performers & Environment**. As fig. 5 makes explicit, the audioscore does not monopolise perception; instead, it frequently redirects listening toward other performers, the acoustic space, or ambient sound. Curved arrows indicate that this listening is continuous and multi-directional: performers listen simultaneously to their own instruction stream, to other singers, and to the surrounding environment.

Listening informs **Embodied Actions**, the next stage in the flowchart. Performers integrate what they hear into situated vocal actions, movements, orientations, and choices of proximity. The audioscore does not prescribe how these decisions should be embodied; performers draw on bodily habit and social competence to realise instructions in context.

These individual actions aggregate into **Emergent Spatial-Social Configurations**, such as dyads, clusters, temporary focal figures, or dispersed parallel activity. These configurations are not choreographed but arise from performers responding at slightly different times and from different spatial positions to related, yet non-identical, instructions. A crucial feature of fig. 5 is the feedback loop from spatial-social configuration back to listening. As performers hear one another and perceive evolving spatial relations, the configuration itself becomes an object of attention that shapes subsequent actions. This recursive process constitutes the above-mentioned procedural social chore-

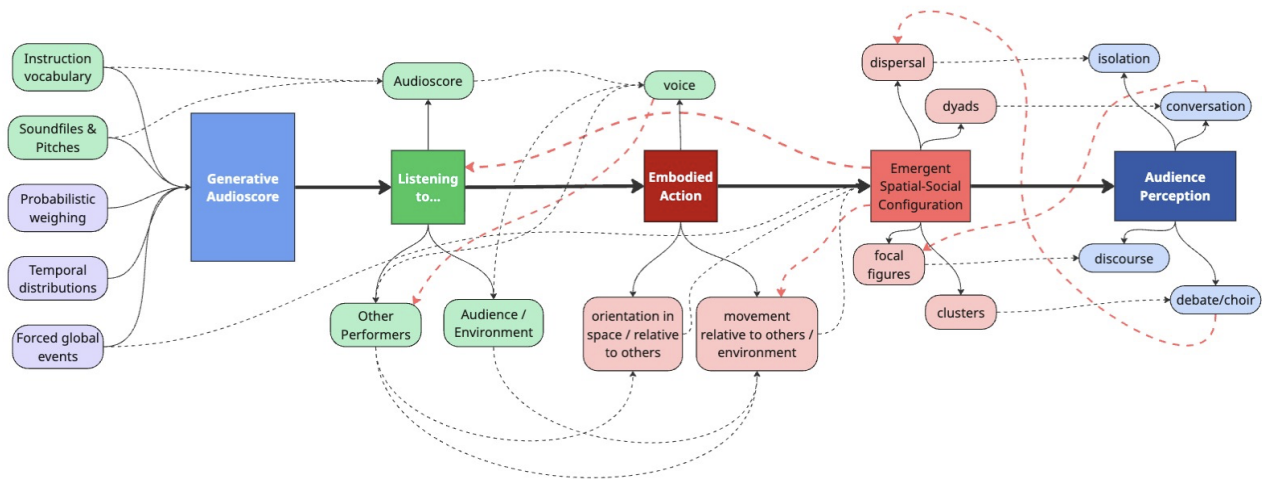


Figure 5. Visualization of the generative audioscore of *Villanelles de Voyelles* as a multi-layered process in which listening, rather than gesture or movement, functions as the primary motor of musical, spatial, and social action. Although organised as a left-to-right flowchart, the diagram includes feedback loops (red arrows) that indicate a recursive rather than linear process

ography: social organisation unfolds through rule-based interaction and continuous listening rather than through predefined movement patterns.

The rightmost cluster, **Audience Perception**, is connected by interpretation rather than causation. Audiences tend to perceive the configurations of people and noises as socially legible situations—conversation, canvassing, deliberation, or collective attention—even though performers are not instructed to represent such scenes. Theatrical meaning thus emerges as a non-representational theatrical affordance of the generative audioscore.

Taken together, the steps visualised in fig. 5 show that the audioscore composes relations of listening rather than gestures or choreography. Sound, space, and social interaction are coordinated through recursive attention and response, allowing theatricality to emerge without representation.

One notable lacuna in this particular instance of the audioscore is the lack of feedback of the performance into the generative audioscore software itself, through the analysis of sensorial or audiovisual live-feeds. This would require live-AI analysis of complex sonic events and of social situations, which was not the state of AI research in 2022³. Future iterations of this and similar works might well integrate such a major feedback loop – which in turn would

significantly change aesthetic affordances and therefore require an entirely new artistic concept.

8. CONCLUSIONS

This paper has proposed the concept of the “*Villanelles de Voyelles*” generative audioscore as a notational paradigm in which compositional intent is embedded in probabilistic, temporal, and relational structures rather than in fixed sequences of events. Through *Villanelles de Voyelles II*, we argue that such scores can compose not only sound and time, but also space and social interaction.

By enabling free movement, elastic coordination, and relational instruction, generative audioscores afford procedural social choreography and non-representational theatrical affordance. Theatrical meaning emerges not because performers represent social roles, but because they navigate structured situations using embodied social intelligence.

For notation research, this suggests that scores can operate as executable social frameworks, composing conditions for interaction rather than prescribing outcomes. Generative audioscores thus may extend the domain of notation beyond symbol and sound, into the realm of theatrically or choreographically enacted social meaning and the aesthetic atmospheres they give rise to.

9. REFERENCES

- [1] S. Bhagwati, “Elaborate audio scores: Concepts, affordances and tools,” in *Proceedings of the International Conference on Technologies for Music Notation and Representation—TENOR*, vol. 18, 2018, pp. 24–32.
- [2] A. Lepecki, *Exhausting dance: Performance and the politics of movement*. Routledge, 2006.
- [3] —, *Singularities: Dance in the age of performance*. Routledge, 2016.

³ By 2022, individual components of such a feedback system existed at a research level—most notably multi-person tracking, pose estimation, audio-visual active speaker localisation, and coarse proxemic grouping from video. However, no mature or robust AI systems were capable of integrating multi-channel audio and multi-view video streams in real time to infer social configurations, attentional relations (e.g. who orients or listens to whom), or transitions between social states in complex, unconstrained and not always uniformly lit stage performance situations. Moreover, existing research prototypes typically required tightly controlled camera geometries, lighting conditions, and calibration procedures, making them impractically complex for a repeatable, portable, and touring-compatible stage setup. As such, closing the feedback loop between live performance and generative audioscore through AI-based audiovisual analysis was neither practically nor aesthetically viable in 2022. Whether subsequent advances in AI research have made such an integration feasible—and how this would transform the theatrical affordances of generative audioscores—constitutes an open question for future research-creation projects.

- [4] A. Hewitt, *Social choreography: Ideology as performance in dance and everyday movement*. Duke University Press, 2005.
- [5] P. Handke, *The Hour We Knew Nothing of Each Other*. New York: Farrar, Straus and Giroux, 1997, originally published as *Die Stunde da wir nichts voneinander wussten*. Frankfurt am Main: Suhrkamp, 1992.
- [6] S. Beckett, "Act without words i and act without words ii," in *The Complete Dramatic Works*. London: Faber and Faber, 1986, act Without Words I originally written in 1956; Act Without Words II originally written in 1957.
- [7] S. Foster, *Choreographing empathy: Kinesthesia in performance*. Routledge, 2010.
- [8] W. Forsythe, "Improvisation technologies: A tool for the analytical dance eye," ZKM — Center for Art and Media Karlsruhe, 2011, originally produced 1994–1996 by ZKM — Center for Art and Media Karlsruhe; currently available as linear video on YouTube. [Online]. Available: <https://www.youtube.com/watch?v=Vx0fe9R1D7E>

INTERPRETING AUDIO SCORES: EXPERIENCES AND INSIGHTS

Barbara Maria Neu

University of Music and Performing Arts
Vienna
neu@mdw.ac.at

ABSTRACT

This paper provides a resuming insight into my experiences of interpreting audio scores as a clarinetist. It draws on moments from my experience that began with my first encounter with audio scores and have since continued through sustained and intensive engagement within the framework of a research project about audio scores. The experiences from the initial Encounter with an audio score through to the Performance are described from the perspective of a clarinetist trained in the classical music tradition, who has repeatedly engaged with contemporary music out of personal interest, without having completed any specialized training in this field. The interpretation of contemporary music, extended techniques and collaboration with composers, usually based on written scores, are therefore familiar to me. However, they do not constitute the everyday content of my work as a musician. The starting position of working with musicians, who are somehow familiar with contemporary music but do not work with it in their everyday life, is, in my view, one that composers in contemporary music frequently encounter and will continue to encounter and one for which audio scores can offer unexpected advantages and possibilities.

1. INTRODUCTION

1.1. Audio Scores: Definition

Before turning to my own experiences, I would like to provide an overview of some methods and approaches to audio scores. S. Bhagwati [1] proposes a possible definition by describing audio scores as scores that primarily use auditory means of communication to convey defined aspects to performers. These aspects can be communicated in different ways and can serve various functions: audio scores may contain information or instructions, encourage imitation and provide examples. Auditory scores can therefore exist entirely on their own or function as a supplement, extension or completion of visual scores, offering advantages and facilitation for performers (see J. Bell [2]). Conversely, visual and/or printed elements of scores may also serve as supplements to audio scores.

1.2. Situative / Live Audio Score, Fixed Audio Score

In principle, a distinction can be made between audio scores that are generated in the moment (for example, generated live by the composer) and thus react to what is happening in the concert and scores that are completed in

advance (for example, fixed media) and played back during the performance (see S. Bhagwati [1] and E. Schimana [3]). Audio scores therefore do not necessarily have to be situational or real-time.

1.3. Acoustic Memory

E. Schimana [3] presents another method in which the score is completely absent during the performance. *Acoustic Memory* describes an approach in which the score is conveyed through audio and memorized by the performers during the rehearsal process. During the performance, it is then presented from memory.

1.4. Different Approaches

Depending on the approach and the composer's instructions, which are discussed during rehearsals, the conveyed information can be handled either more strictly (accurately reproducing only what is heard) or more intuitively (treating the information as guidance for further development). These agreements have a significant influence on the interpretation. Even in the former case, the individual and subjective interpretations of the musicians may vary and thus play a decisive role in the final outcome. In the latter case, the scope of interpretation can extend so far that it may be understood merely as an impulse for improvisation.

1.5. Audible Score as an Artistic Medium

An interesting aspect of audio scores is the possibility of the score functioning as an artistic medium. Depending on the mode of transmission, the score may be delivered via headphones or loudspeakers (or other experimental means). This allows composers, depending on their artistic intentions, to present the score audibly as part of the overall work (provided it is played back live). If the score is played through loudspeakers, it is the composer's decision whether, for example, the entire score (all tracks) is played clearly through four loudspeakers in the space or whether each musician has their own loudspeaker (and individual track) placed nearby, possibly audible only to them. Hybrid forms are also possible.

2. Experiences in Live Generated Audio Scores

One of my first experiences with audio scores was with live-generated audio scores. I worked with a composer who already had several years of experience with various musical ensembles and has applied audio scores as a

long-standing compositional method. During the process, we carried out, among other things, small performances. In addition, there was a recording process primarily intended for psychoacoustic analysis. As a result, the following presents subjective analyses and observations from the perspective of the performer, that is, from my perspective.

2.1. Setting

Even for specialized terms such as live-generated audio score, there are numerous interpretations and variations. The setting I will describe here is as follows: each musician positions themselves next to a loudspeaker. The composer sits at her mixing console, which she operates live. The score is transmitted through the loudspeakers in separate tracks (one track per musician), with the goal of as precise an imitation as possible. What is particularly interesting is that the score is generated live and the composer reacts directly to what I, as the performer, play.

2.2. Development, Methods and Rehearsals

Before the first actual rehearsal in the final setting, there was a sort of preliminary planning sessions with the composer, still without the electronic setup. During this session, the fundamental basics of the composition were clarified, primarily through the composer's questions (possible pitch range, possible extended pitch range, etc.). In addition, four different "Performance methods of the score" were discussed: 1. Long tones, 2. Two tones simultaneously, 3. Noise, 4. Staccato.

2.2.1. Methods: Single Long Tones

During the first practical rehearsal in the electronic setup with loudspeakers, we were able to start without further preparation. Crucial for me was the prior information that only three tones would appear in the first section. My expectation - or rather fear - that I would have to "guess" the notes in audio scores was fortunately dispelled. Instead, the live-generated audio score conveyed to me how long and how loud the tones of the score, that is, one of the three agreed-upon tones, should be played. The function of the audio score, therefore, is not to guess pitches but to provide impulses and instructions regarding the duration, intensity and sequence of the tones.

2.2.2. Methods: Two Tones Simultaneously

In retrospect, the long tones almost functioned as a preparation for the second method. Without explicitly discussing it, during the course of practicing the long tones, suddenly not just one tone but two tones simultaneously were transmitted through the loudspeaker, that is, through the score. Perhaps it was the previous instruction to play as precisely as possible what is heard that led me to stop attempting to play single tones. Rather, there was a kind of "falling into" the loudspeaker. When two tones were played simultaneously, this did not result in playing one tone or the other but in an attempt to imitate the simultaneous sound of both tones. Technically, this involved a combination of partially opening or closing keys, adjusting the (lip) embouchure and possibly changing the direction of air and breath, which produced this attempt at imitation. On the clarinet, alternative fingerings are quite common when playing quarter tones.

However, these are rarely, or only in very specific personal agreements, combined with the embouchure and air adjustments I employed here. Each moment is not fully reproducible. Whenever the combination of the two tones in the score repeated, the attempt at imitation triggered a different clarinetistic playing combination.

2.2.3. Methods: Noise

Similar to the second method, the third method developed in a comparable way. Without further technical arrangements, a kind of noise or flutter was added to the live-generated score alongside the long tones. The most obvious way to imitate such sounds on the clarinet is the flutter-tonguing technique, which I initially applied in the low register. A defining aspect of the entire rehearsal process across all ranges and methods was the ongoing, direct exchange with the composer. Through these discussions, it became clearer in which direction the composition should proceed.

The flutter-tonguing method worked wonderfully in the first two ranges. However, the combination of extreme pianissimo and flutter-tonguing could not be executed technically in the highest register of the instrument. Therefore, I switched to a method that primarily consisted of producing a lot of air in the tone (i.e., noise).

2.2.4. Methods: Staccato

In comparison to the other three methods, the fourth approach is somewhat distinctive. At the beginning of this stage, individual tones were again transmitted, this time short, that is, staccato. Initially, these were single tones of the same pitch. Subsequently, a second and third tone were introduced, presented in quick succession and eventually even simultaneously. In addition to the tones of the lower range, occasional tones from the second range were incorporated. As in the previous methods, working in the mode of "falling into" the loudspeaker resulted in an interpretation characterized by a rapid sequence of staccato tones across varying pitches, as I aimed to "capture" as many tones as possible. Notably, compared to a traditional written score, achieving such an energetic, fast and spontaneous outcome would be challenging, or only feasible with extensive rehearsal and practice.

Regarding the technical implementation, the composer indicated that for this method, a loop track was created for each pitch, causing each tone to be repeated at regular intervals (e.g., every 4.78 seconds). Additional tones were assigned separate tracks with differing temporal spacing, resulting in a score in which tones are sometimes presented simultaneously but more often sequentially and in varying order.

2.3. Merging with the Score and the Concept of Mimesis

As mentioned above, this process involves a kind of "falling into" the loudspeaker, which I would like to examine in more detail here. In art theory, the concept of *mimesis* has repeatedly played a central role. W. Benjamin [4] describes a situation of observing and hunting a butterfly and entering a state in which the attempt at imitation goes so far that the butterfly is not merely imitated, but a sensation arises of almost becoming one with the butterfly. What is at stake here is a

form of complete immersion or merging, perhaps a meditative state, in which technical processes recede from conscious awareness. Already Plato [5], in his work *Politeia*, understands mimesis as the imitation of natural processes or of nature itself.

I am able to draw on this state and this experience in my encounters with live-generated audio scores. After a certain period of rehearsal, at a later stage in which verbal communication may have become less prominent, I had the impression of merging with the loudspeaker, or more precisely, with the sounds of the score itself. It felt like an invisible connection: the loudspeaker thus became a playing partner or ensemble colleague and rather than experiencing the situation as following its instructions, I increasingly perceived it as making music together. In some situations, the setup was arranged in such a way that I was physically turned away from the composer and therefore unable to see her. I believe that this further intensified the focus on what was heard and deepened the sense of merging with the loudspeaker. The recordings confirmed that at certain points the score and the clarinet could no longer be clearly distinguished temporally. At some point, I was playing almost simultaneously with the score, without actually knowing when the next tone would occur.

2.4. Requirements and Precision

In the course of the rehearsal and analysis process, I repeatedly arrived at a particular realization: strictly speaking, I am also unable to fulfill the requirements of the audio score with complete precision. Even when, based on a state of merging informed by the concept of mimesis, I come very close to the score, it is impossible to perform it in exact accordance with it. In this context, however, I also question how closely the intentions of composers can be realized when performing from a written score.

In follow-up discussions with the composer, it became clear that the instruction to play as precisely as possible with the score functions more as a necessary means to an end than as an absolute requirement for the interpretation. As mentioned above, this attempt induces a meditative, highly concentrated state that results in a sense of merging with the score and consequently in a resonance with its musical energy. In this form of audio score, it is ultimately not decisive whether my clarinet tone ends in perfect temporal alignment with the score as the tone will become very soft and may "disappear" behind the potentially longer-lasting score. For listeners, such discrepancies would hardly be perceptible. Moreover, it became clear to me that I also have the option of omitting certain tones of the score while playing others. An option that arises, at least in part, from the technical possibilities described above.

2.5. Interpretation and Interaction

Where, then, does the scope for interpretation lie within this form of audio score? Naturally, the interpretative freedom is more limited than in scores that function primarily as impulses. Nevertheless, certain parameters remain embedded within the score, as it is still the performers' interpretation that determines, for example, how dynamic indications are realized in practice.

Decisions regarding which tones are played and which are omitted may likewise be understood as interpretative acts. However, the implementation of the methods described above is of central importance for the interpretation. The manner in which noise or flutter is produced, or how the score's impulse to play two tones simultaneously is realized, varies from performer to performer.

A crucial aspect of interpretation in this context is what I would describe as the "interaction between performer and composer". The live generation of the audio score does not only mean that the composer directly influences what the musicians play. The reverse is equally true. My interpretation as a performer, in turn, influences the composer and thus the score itself. The composer hears what I play and responds accordingly, for instance by retransmitting the signal at a different dynamic level - more reduced and soloistic, or more intense. As a musician, I therefore exert an indirect influence on the score, even though this aspect does not involve conscious decisions to "manipulate" the score.

2.6. Another Experience of a Situational Audio Score

At this point, I would like to mention another process involving a different composer and a different approach to situational audio scores. The compositional concept likewise falls within the category of situational audio scores. However, in this case, the score is not transmitted via live-generated electronics. Instead, the rehearsal process involved two musicians (clarinet and double bass) positioned facing one another, both instructed to play the same pitch. The central instruction was to listen closely to the resulting acoustic beats and resonances that emerge when identical pitches are produced and either to work against these phenomena by reducing them or reinforce them.

This rehearsal took place at an early stage of the project. At a later point, the approach is intended to be realized within an ensemble context, involving two distinct groups. Even at this preliminary stage, however, it can be observed that a form of interaction emerges here as well between the "score", in this case, the other musician and myself as the performer. The score does not originate from an external source, but is instead produced by the performers themselves in the moment, that is, situationally.

3. Experiences in Acoustic Memory

As described above, the method of *Acoustic Memory* refers to audio scores that are performed live from memory during the performance. I had the opportunity to explore this method as well in collaboration with a composer who has repeatedly integrated it as a consistent and established element of her artistic practice.

3.1. Setting und Source Material

During the initial meeting, I received the audio score in the form of a 45-second audio file. This file consisted of relatively short, rhythmic, abstract electroacoustic sounds, characterized by their typical timbres and a pronounced noise-like quality. The conceptual framework of the work was such that each ensemble member would receive a (different) audio score of this type for each

movement of the piece. In addition, a temporal structure was defined that specified when these audio scores were to be performed.

On stage, each musician is therefore provided with a clock and a timeline indicating at which second the memorized 45-second insertion should be performed. This insertion occurs multiple times over the course of each movement.

3.2. Translation and Implied Interpretation

Reproducing melodies and sounds from other instruments is, of course, not unfamiliar to me and translating "noise" onto my instrument has also become familiar through various encounters with improvisation and contemporary music. Nevertheless, finding a melody based purely on electroacoustic sine tones and translating it onto the clarinet can be described as challenging.

At home, my initial effort focuses on gradually familiarizing myself with the file. I listen to it several times in succession. The tones are high and almost piercing. Some sounds follow so quickly that I wonder how they could ever be executed on my instrument purely from a technical standpoint. Were it fully notated, it would likely require hours of practice. As I have experienced with classical works, when something seems unmanageable, a structured approach is required to progress step by step. I therefore divide the piece into segments: ten seconds, another ten, a total of four ten-second segments, plus an additional five seconds at the end. The next step consists of gradually translating the piece, always in ten-second increments, for myself and my instrument. In retrospect, I can say that this is where my personal interpretation becomes decisive. Another musician might translate the piece differently, perceiving and rendering the sounds in another way. My artistic contribution thus represents a substantial part of the final result.

I begin with the first ten seconds, translating them onto my clarinet and recording them, thereby creating another personal audio score. I select a few high tones that start softly and gradually become louder, add anticipations and incorporate some multiphonics and slap-tongue techniques. This process allows me to make progress more quickly than expected. I gradually approach the audio file, become more familiar with it, discover a certain internal rhythm and coherence and perhaps begin to feel some affinity for sounds that initially seemed so foreign. The process is similar to learning classical music and written scores: I recall first studying the Brahms Clarinet Sonata at nineteen, when I had to play it repeatedly to follow the full arc of the work effectively.

3.3. Rehearsal with the Composer

After one week, the composer and I met again. She pointed out a few components that were missing in my interpretation, particularly some of the "crackling" sounds. In general, I had interpreted the piece slightly too loudly and the tones lacked sharpness at their onset and offset.

We discussed how precisely the percussive crackling at the specific points in the file needed to be executed, a point that had initially caused me some concern. However, it became clear that the fundamental presence

of the percussive elements as a counterpoint was what mattered most. Once again, the goal was to capture the style of the score. During this process, I reflected that in written scores, considerable time would likely be spent on technically perfect execution. I began to wonder whether conveying the style and the general intent of the composition might be far more important.

After a few attempts, I played somewhat more softly, sharpened the tones and emphasized certain points with percussive finger noises on the keys and slap-tongue techniques. The composer expressed satisfaction and I gradually began to understand the magic inherent in audio scores.

3.4. Presentation

Compared to other types of audio scores, the approach using *Acoustic Memory* is relatively well-prepared for instrumentalists. The presentation took place within the context of a workshop, followed by a discussion. After the presentation, we discussed this approach and it quickly became evident and self-evident that, when listening as an audience, it is not discernible that the work is based on an audio score. It could just as well have been performed from a written (or another type of) score.

4. Further Observations and Conclusion

4.1. The Concept of Mimesis

As discussed above, I would like to explicitly refer once again to the concept of mimesis, which, in my experience, is relevant across all the audio score methods described here. I find it particularly interesting to consider how and in what form the concept of mimesis, understood as imitation and emulation, applies to written scores. In my classical instrumental training, I experienced that mastering the technical skills on the instrument and thus the ability to play the notes accurately, occupies a significant amount of attention, often pushing the methods of mimesis into the background. This is an aspect that manifests differently in the context of audio scores.

4.2. Interpersonal Aspects and the Question of Hierarchy

Throughout the various development processes, which included rehearsals as well as concerts, presentations and workshops, the question of hierarchy arose repeatedly. At first glance, the setting of live-generated audio scores might appear to be characterized by clearly defined hierarchical structures. Based on my experience, however, I would clearly challenge this assumption. Due to the close collaboration with the composer and the interaction described above, I would argue that my personal interpretative scope and my influence on the composition were significantly greater than in the case of written scores. Similarly, within the method of *Acoustic Memory*, I experienced a broader interpretative space and a higher degree of decision-making freedom than I typically encounter when working with written scores.

4.3. Ease of Playability and Practice

In my view, one of the most significant observations concerns the considerably greater ease with which audio scores can be realized in comparison to written scores. In interpreting audio scores, I encountered playing techniques that would likely require extensive and prolonged practice and that, moreover, could certainly not be notated in the same way. In my experience, composers are more concerned with conveying a particular atmosphere and energy. Elements that may recede into the background during the practice and technical realization of written scores. Audio scores therefore offer a much more accessible mode of engagement than written scores, even if the absolute precision of written notation, down to every detail, is potentially relinquished. The decision to employ audio scores thus ultimately lies with the composer, depending on the specific requirements and artistic intentions of the respective work.

Acknowledgments

This research was funded in whole or in part by the Austrian Science Fund (FWF) [grant DOI 10.55776/AR824].

The paper is part of the project "Sound as Score". Project management: Elisabeth Schimana - University of Music and Performing Arts (mdw)/Artistic Research Center (ARC). Project partner: Susanne Kogler - University of Graz (KFU)/Department of Arts and Musicology and Piotr Majdak - Austrian Academy of Sciences (ÖAW)/Acoustics Research Institute (ARI)

I would like to express my sincere thanks to Elisabeth Schimana for the collaboration on working with audio scores. I am also grateful to Katharina Pollack, Piotr Majdak for the collaboration in the field of psychoacoustics and Peter Venus for producing the sound recordings. Further thanks go to Susanne Kogler and Nicolas Müller-Lorenz. I would also like to thank the FWF and the ISF at the Austrian Academy of Sciences (ÖAW), where the recordings were made.

4. REFERENCES

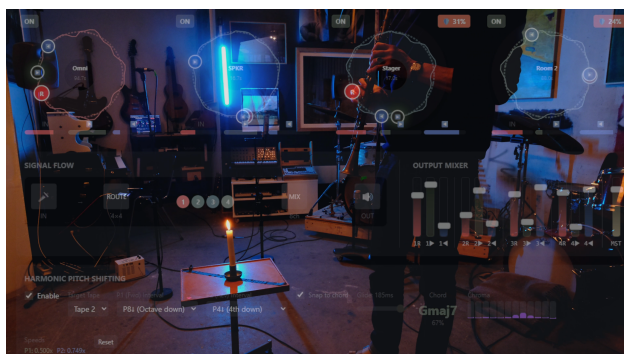
1. Sandeep Bhagwati, *Elaborate Audio Scores: Concepts, Affordances and Tools*, Proceedings of the International Conference on Technologies for Music Notation and Representation - TENOR'18
2. Jonathan Bell, *Audio-Scores. A resource for Composition and Computer-Aided Performance*, Final Submission for Guildhall School of Music & Drama, 2016
3. Elisabeth Schimana, *Sound As Score: Live Generated Audio Score and Audio Score Based on Acoustic Memory*, Proceedings of the International Conference on Technologies for Music Notation and Representation - TENOR'19
4. Walter Benjamin, *Berliner Kindheit um neunzehnhundert*, Projekt Gutenberg-DE, posthum 1950; im Suhrkamp Verlag Berlin 2010
5. Platon, *Der Staat X. Buch*, 375 V. Chr.; Projekt Gutenberg-DE

PLAYING THE ROOM: TAPEPLAYER AND THE IMPROVISED AUDIO SCORE

Jonas Sjøvaag

University of Agder, Norway
jonas@jonassjovaag.no

ABSTRACT



This paper presents TapePlayer, a tape echo simulator designed as a score-framework for improvising musicians. Drawing on the incommensurate loop techniques of Brian Eno's ambient works, Terry Riley's time-lag accumulators, and Alvin Lucier's room-as-instrument approach, TapePlayer provides a compositional infrastructure: four independent delay lines with configurable lengths, bidirectional playback heads, and—crucially—the acoustic room itself as primary feedback network. The score is not the sound that emerges but the framework that shapes it: system configuration, microphone placement, and spatial acoustics. The improviser activates this structure, generating material that the system transforms and returns. Rather than prescribing pitches or rhythms, TapePlayer establishes conditions for mimetic engagement—performers identify with, respond to, and transform their own sound as it returns altered by room and machine. I examine this relationship through the concept of mimesis as theorized by Paddison: not imitation-as-copying but embodied identification with sound. The paper discusses TapePlayer's technical architecture, its application as score-script for improvisation, and the psychoacoustic implications of a framework that never produces the same result twice. A video demonstration illustrates the mimetic relationship between performer, system, and space.

Copyright: © 2026 Jonas Sjøvaag. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

What does it mean to "read" a sound? When a performer encounters an audio score—a sonic environment intended to guide or provoke musical response—they engage in an act fundamentally different from reading traditional notation. There is no prescribed pitch, no fixed rhythm, no determinate instruction. Instead, there is sound itself, and the performer must decide how to relate to it.

TapePlayer is a tape echo simulator that I have developed as a Python script, and it acts as a response generator that provides an improvisational framework situated in a specific room. It also encourages improvisational musicians to see its response, or feedback if you will, as audio scores evolving in real time, based on the input from the player(s). The system creates continuously evolving sonic textures that performers interpret and respond to during improvised performance. My central argument is that this interpretive act is best understood through the concept of mimesis—not in the narrow sense of imitation or copying, but in the broader sense articulated by Paddison [1] and others: mimesis as embodied identification, as a mode of "identifying with" rather than "imitation of."

The TENOR 2026 conference theme, "Sound as Score," invites consideration of how audio materials function as notation—how they guide, constrain, and enable musical action. The associated research project asks how the concept of mimesis relates to the aesthetic analysis and evaluation of audio scores. TapePlayer provides a concrete case study for examining these questions. As a system that generates scores which literally never repeat, it foregrounds the temporal and relational dimensions of mimetic response that remain implicit in fixed audio scores.

While tape-based delay systems have a rich history, the scripting approach the system builds on, its various technical features (outlined in section 3) and the room-as-feedback-network within a mimesis framework for improvisation, appear under-documented in existing literature.

The paper proceeds as follows: Section 2 situates TapePlayer within the historical lineage of tape-based generative systems and introduces the relevant conception of mimesis. Section 3 describes the technical architecture of TapePlayer. Section 4 analyzes how the system functions as an audio score and how performers engage with it mimetically. Section 5 describes the accompanying

demonstration. Section 6 concludes with implications for audio score composition and research.

2. BACKGROUND

2.1 Tape Loops and Generative Systems

TapePlayer emerges from a rich lineage of tape-based generative music systems. Terry Riley's time-lag accumulator (1963) established the foundational technique: two tape machines configured so that sound recorded on one is played back on the other after a delay, with feedback allowing material to accumulate and transform over time. Robert Fripp and Brian Eno refined this approach in their Frippertronics system (1972), using it to create dense, slowly evolving guitar ambiances [7].

Eno's subsequent work, particularly *Music for Airports* (1978), introduced a crucial innovation: incommensurate loop lengths. Rather than using loops of equal or simply-related durations, Eno employed loops whose lengths could not be expressed as simple ratios of one another. As he described it, loops of $23\frac{1}{2}$, $25\frac{7}{8}$, and $29\frac{15}{16}$ seconds would not return to their starting alignment for approximately 27 days [2]. The result is music that, while constructed from repeating elements, effectively never repeats at the macro level.

Steve Reich's phase music offers a related but distinct approach. In works like *Piano Phase* (1967), two performers begin in unison and gradually shift out of phase, creating emergent patterns from their temporal displacement [3]. Where Eno's incommensurate loops create non-repetition through mathematical irrationality, Reich's phasing creates cyclical patterns of convergence and divergence.

TapePlayer synthesizes these approaches. It employs multiple delay lines with configurable lengths that can be set to incommensurate values (following Eno), bidirectional playback that creates phase-like relationships between forward and reversed material, and—most significantly—an architecture designed to use the acoustic room itself as the primary feedback network.

This last aspect connects TapePlayer to another lineage: Alvin Lucier's *I Am Sitting in a Room* (1969), where the acoustic properties of a space gradually transform and eventually consume the source material [4]. Where Riley, Eno, and Fripp used tape machines as closed feedback systems, Lucier opened the loop to include architectural space. TapePlayer follows Lucier in treating the room as an active participant in the sound-generating process.

Recent artistic research has examined human-machine improvisation through archiving and playback systems. The Norwegian project *Goodbye Intuition* (Grydeland, Neumann, Qvenild, 2017-2020) developed Kim-Auto, a Max/MSP system that records performer input, archives

it, and plays back transformed fragments—cut up, reshuffled, time-stretched, reversed [8]. They describe its output as a "turbulent mirror," producing "blurred mimesis" of the performer's material. Their aim was to create an autonomous improvisational partner with its own "personality." TapePlayer shares the tape-loop lineage but differs in framing and architecture: where Kim-Auto aspires to be a co-improviser, TapePlayer positions itself as a score-framework. The system makes no decisions about when or what to play back—the temporal relationships are fixed by loop lengths, and the room's acoustics provide the feedback network. The mimetic relationship is therefore inverted: rather than the machine producing blurred mimesis of the performer, the performer practices mimesis toward their own transformed material as it returns through space and time.

2.2 Mimesis and Audio Scores

The concept of mimesis has a long and contested history in aesthetics. From Plato's understanding of art as imitation twice removed from ideal forms, through Aristotle's more positive valuation of mimesis as creative representation, to modernist rejections of imitation in favor of abstraction, the term has carried shifting meanings and valuations.

In the context of music, mimesis has often been associated with programmatic representation—birdsong in Messiaen, thunder in Beethoven's Sixth Symphony. Allen S. Weiss's *Varieties of Audio Mimesis* [5] expands this understanding, proposing a typology of imitative strategies in sound art that encompasses not only direct sonic imitation but also evocation, stylization, and abstract correspondence.

More relevant to the present discussion is Paddison's [1] Adornian conception of mimesis as embodied impulse. Drawing on anthropological theories of mimicry (particularly Caillois [6] and Benjamin), Paddison argues that mimesis in music is better understood as "identifying with" than as "imitation of." Musical mimesis is not primarily a matter of representing external objects or phenomena; it is a mode of relation, an embodied orientation toward sonic material.

This conception proves particularly apt for understanding how performers engage with audio scores. When a musician responds to an audio score, they are not typically attempting to copy or reproduce what they hear. Rather, they are entering into relation with it—identifying patterns, sensing trajectories, finding points of connection and divergence. The response may involve imitation in a local sense (echoing a gesture, matching a timbre), but the overall orientation is one of identification and relation rather than reproduction.

The Sound as Score research project asks how mimesis relates to the aesthetic analysis and evaluation of audio scores. I propose that audio scores can be understood as environments for mimetic engagement—sonic situations

that invite and structure the performer's identificatory response. The quality of an audio score, on this view, depends not on how well it specifies a predetermined outcome but on how richly it affords mimetic engagement.

3. TAPEPLAYER: TECHNICAL ARCHITECTURE

TapePlayer is implemented as a standalone Python application. The system comprises up to four independent tape loops, each with distinct record and playback mechanisms, connected through a routing matrix and mixing infrastructure. Additionally, audio hardware with mic preamps is an obvious necessity, and to make the software easy to work with, it can work with any number of inputs from 1 to 4.

3.1 Core Architecture

The signal flow routes four inputs through a routing matrix to four tape loops, each containing a record head, forward and reverse playback heads, a variable-length buffer (2-120 seconds), and an independent feedback path. Eight playback outputs feed into a mixer producing stereo output. All of this happens internally in the Reaper DAW, so the script itself uses ReaRoute 1 - 4 for mono outputs and ReaRoute 5+6 for stereo inputs. This way, any kind of hardware can be used with it, there's no particularity or tie-in with any piece of gear.

Inside the software, each of the four inputs can be routed to any combination of the four tape loops. Each loop maintains an independent circular buffer with a configurable length ranging from 2 to 120 seconds. Loop lengths can be set individually, randomized within a specified range, or recalled as part of saved presets. In my usage, these presets can be seen as compositions. Each preset provides an improvisational framework that differs from the others, affecting the players and how they relate to the audio, precisely like a sound score.

This configurability shifts the question of temporal relationships from technical specification to compositional decision. The system does not enforce incommensurate ratios; rather, it provides the infrastructure for a composer or performer to design temporal relationships according to their aesthetic intentions. A preset might specify four loops of 7, 11, 13, and 17 seconds—prime numbers ensuring long recurrence cycles—or four loops of 4, 8, 12, and 16 seconds for more periodic, pulse-aligned relationships. The choice is compositional: what kind of temporal environment should this audio score create?

In my own practice, I typically work with 2x2 connected loop lengths, two very long (100 seconds) and two shorter, composing with them using playback direction, speed and volume. TapePlayer as a system remains agnostic, offering temporal architecture as a parameter space for compositional exploration.

3.2 Bidirectional Playback

Each tape loop features two playback heads: one reading forward through the buffer at normal speed, one reading backward. These heads operate independently, creating a continuous counterpoint between the original temporal direction of recorded material and its reversal.

The reversed playback head introduces a particular kind of mimetic complexity. Reversed sound retains timbral and dynamic characteristics while inverting temporal envelopes—attacks become decays, decays become attacks. The result is material that is simultaneously recognizable and transformed, familiar and strange. Performers responding to TapePlayer's output must navigate this dual temporality, identifying with material that is always already doubled.

3.3 Tape Modeling

To evoke the character of analog tape systems, TapePlayer incorporates several modeling elements: saturation (soft clipping at the record stage creating harmonic coloration), hiss (low-level noise simulating tape noise floor), wow and flutter (subtle pitch modulation simulating mechanical irregularities), and high-frequency rolloff (progressive attenuation with each feedback pass).

3.4 Chroma-Based Pitch Analysis

TapePlayer incorporates real-time harmonic analysis to inform pitch-shifting decisions on selected tape channels. A chroma analyzer extracts the predominant pitch class from incoming audio, representing harmonic content as a twelve-element vector corresponding to the chromatic scale. This analysis then guides automatic pitch-shifting of a designated tape channel, transposing playback by selectable intervals relative to the detected root.

This feature enables harmonically responsive behavior without requiring the system to generate new material. If the performer plays in A minor, the selected tape channel can be shifted to produce material a third or fifth above the detected root, creating harmonic reinforcement or tension depending on the interval chosen. The performer thus hears their own material returned not only temporally displaced and spatially transformed, but harmonically recontextualized.

To my knowledge, no existing delay or looper system uses chroma analysis to drive pitch-shifting in this manner. Commercial key-detection tools exist for informational purposes, but dynamic harmonic modulation of delay parameters based on real-time tonal analysis appears to be an interesting and new addition this particular script can offer.

3.5 The Room as Feedback System

While TapePlayer includes internal feedback paths for each loop, the system was fundamentally designed to *play the room, with the musician*. Rather than relying primarily on digital feedback routing, TapePlayer treats the acoustic space itself as the feedback network. The system outputs to loudspeakers; microphones placed in the room capture not only the performer but also the room's transformation of TapePlayer's output; this captured sound feeds back into the system.

This design philosophy connects TapePlayer to Alvin Lucier's *I Am Sitting in a Room* [4], where successive recordings through a space gradually reveal and amplify the room's resonant frequencies until speech dissolves into pure resonance. TapePlayer operates on a similar principle but in real time and with continuous performer involvement. The room's acoustic character—its reflections, absorptions, resonances—becomes part of the score-generating mechanism. Different spaces produce different TapePlayer behaviors, and the composition will dissolve only by choice of the musician; the minute the instrument(s) stops playing, the composition will start to deteriorate towards an end, being in a room, until it dissolves into nothing.

Microphone placement therefore becomes a compositional parameter as significant as loop length. Close miking emphasizes the performer's direct input; distant miking brings more room coloration into the system; strategic placement near reflective surfaces can create specific resonant feedback paths. The performer navigates not only TapePlayer's temporal streams but also their own sound returning transformed by architectural acoustics.

This approach resonates with an installation work I have presented earlier, where a system using speech-to-text and text-to-speech created a feedback loop between human speech and machine interpretation. As the piece progressed, transcription errors accumulated, the machine's "understanding" drifted further from the original utterance, and language gradually dissolved into abstraction. TapePlayer enacts an analogous process with musical material: the room and the system together transform input until its origin becomes indistinct, until the score consists of sonic material whose source—performer, playback, room reflection—can no longer be clearly determined.

This indeterminacy of source intensifies the mimetic situation. The performer identifies with and responds to material that may be their own transformed contribution, the room's resonant commentary, or TapePlayer's temporal manipulations. The boundaries between self and other, between playing and being played, become productively unclear.

4. TAPEPLAYER AS AUDIO SCORE

4.1 Non-Prescriptive Notation

Traditional notation prescribes: it specifies (with varying degrees of precision) what sounds should occur and when. Audio scores, by contrast, present: they offer sonic material without explicit instruction for response. The performer must determine their own mode of engagement.

TapePlayer extends this non-prescriptive quality through temporal indeterminacy. A fixed audio score, however complex, eventually repeats. A performer who has worked with the same fixed score multiple times will develop knowledge of its structure, anticipating what comes next and planning their responses accordingly. This is not necessarily problematic—much valuable music-making involves such preparation and anticipation—but it does represent a particular mode of engagement.

TapePlayer, when configured with incommensurate loop lengths, ensures that the system never presents the same configuration twice. The performer cannot know exactly what comes next because much of what comes next is genuinely undetermined until it occurs in a specific room, based on the specific improvisational output from the instrument. This places the performer in a continuous present, requiring ongoing attention and response rather than prepared reaction.

4.2 Mimetic Modes

I have identified several modes of mimetic engagement that performers adopt when working with TapePlayer:

Timbral identification: The performer matches or complements the timbral qualities of TapePlayer's output—adopting similar tone colors, reinforcing or contrasting harmonic spectra.

Gestural identification: The performer responds to the shapes and trajectories present in the audio score—rising and falling contours, accelerating or decelerating rhythms, expanding or contracting densities.

Temporal identification: The performer aligns with one of the system's multiple temporal streams—locking to the pulse of a particular loop, following the slower evolution of the overall texture, or navigating between temporal scales.

Structural identification: Over longer durations, the performer develops a sense of the system's behavior and responds to emergent structural patterns—the gradual accumulation of density, the periodic emergence and submergence of particular materials.

Contrastive identification: Not all mimetic response involves similarity. Performers may identify what the audio score lacks and provide it—introducing silence into density, sharp attacks into smooth textures, pitched material into noise, or even filling up the loop buffers with one kind of input, then moving on to another, adding a layer of call / response, soloist / accompaniment, and so on.

These modes are not mutually exclusive; performers typically move fluidly among them, and skilled improvisers often engage multiple modes simultaneously. In any case it pays to be deliberate: material committed to the loops will circulate many times before finally decaying, so early choices shape the entire arc of the improvisation.

4.3 The Score That Scores

There is a productive ambiguity in the word "score." As a noun, it denotes a notational document. As a verb, it describes the act of creating such a document, but also the act of marking, inscribing, leaving traces.

TapePlayer scores in both senses. It produces an ongoing audio score that guides performer response. But it also *is scored* by the performer: their input is recorded, transformed, and returned as part of the ongoing score. The distinction between score and performance, between notation and realization, becomes recursive. The performer reads the score and writes it simultaneously.

This recursivity intensifies the mimetic relationship. The performer identifies not only with the system's output but with their own transformed contributions returning to them. They hear themselves as other—displaced in time, reversed in direction, migrated across loops. The mimetic act becomes reflexive: identifying with one's own alterity.

5. DEMONSTRATION

The demonstration component of this presentation will involve showing a recording of a live performance with TapePlayer functioning as audio score. I would prefer to do it on-site, but as it requires a rather elaborate multi-input setup, a suitable room and so on, it's safer and easier to use video.

5.1 Setup

The demonstration will show how TapePlayer is running on a computer with audio output to the room's sound system; microphones positioned to capture both performer and room reflections, feeding back into TapePlayer; one or two performers on acoustic or amplified instruments; and a visual display showing TapePlayer's interface, including real-time visualization of loop contents, playback head positions, and feedback levels.

The specific microphone placement will be determined in response to the instrumentation, the number of musicians involved, and the space's acoustic characteristics, to ensure the room itself becomes an active and optimized element of the audio score system.

5.2 Structure

The demonstration (before showing the video) will proceed in three phases: (1) System introduction (3

minutes)—briefly demonstrating TapePlayer's basic functions and explaining how loop length configuration shapes the system's temporal behavior; (2) Performance (video, 3 minutes)—improvisation with TapePlayer as audio score, seeded with initial material then left to evolve; (3) Reflection (5 minutes)—on the mimetic strategies employed.

6. CONCLUSION

TapePlayer offers a case study in score-frameworks and the mimetic relationships they afford. By generating continuously evolving sonic environments through configurable loop lengths, bidirectional playback, and—most distinctively—acoustic feedback through the room itself, the system creates conditions for sustained mimetic engagement that differs qualitatively from engagement with fixed audio scores or purely digital feedback systems.

The conception of mimesis as embodied identification rather than imitation proves particularly useful for understanding how performers relate to such systems. Performers are not asked to copy TapePlayer's output; rather, they should identify with it, entering into relation with its temporal streams, timbral qualities, and structural trajectories. When the room functions as part of the feedback network, performers must also identify with architectural acoustics—with sound that is simultaneously theirs, the system's, and the space's. This identification is active, selective, and creative—a mode of musical thinking enacted in real time.

Several questions warrant further investigation. How do different acoustic spaces shape the mimetic relationship between performer and system? How does extended familiarity with a score-framework system change the mimetic relationship? How might audio score systems be designed to afford particular mimetic modes while remaining genuinely open?

TapePlayer itself continues to develop. Current work involves integration with machine learning systems for audio analysis, enabling more structured responses to performer input. This raises questions about mimesis in human-machine musical interaction that extend beyond the present paper. More broadly, the accessibility of scripted audio systems has changed: features like automatic feedback ducking, once requiring specialized hardware or commercial software, can now be implemented by any musician comfortable with Python or a similar scripting language. This democratization of situated improvisation tools—systems designed for specific rooms, specific performers, specific compositional intentions—is itself a development worth noting in the context of audio score practice.

The audio score, I suggest, is not merely a notational format but a relational proposition. It proposes a way of

being-with-sound that the performer accepts, modifies, or refuses. TapePlayer, as a system that generates such propositions continuously and unpredictably—and that includes the acoustic room as part of its generative mechanism—foregrounds the relational, spatial, and temporal dimensions of this proposition. To improvise with TapePlayer is to practice mimesis in its fullest sense: not copying, but identifying; not reproducing, but relating; not playing in a room, but playing *with* the room.

7. REFERENCES

- [1] M. Paddison, "Mimesis and the aesthetics of musical expression," *Music Analysis*, vol. 29, no. 1-3, pp. 126-148, 2010.
- [2] B. Eno, *A Year with Swollen Appendices*. London: Faber & Faber, 1996.
- [3] S. Reich, *Writings on Music, 1965-2000*. Oxford: Oxford University Press, 2002.
- [4] A. Lucier, *I Am Sitting in a Room* [Sound recording]. New York: Lovely Music, 1995.
- [5] A. S. Weiss, *Varieties of Audio Mimesis: Musical Evocations of Landscape*. Los Angeles: Errant Bodies Press, 2008.
- [6] R. Caillois, "Mimicry and legendary psychasthenia," *October*, vol. 31, pp. 16-32, 1984.
- [7] E. Tamm, *Brian Eno: His Music and the Vertical Color of Sound*. New York: Da Capo Press, 1995.
- [8] I. Grydeland, A. Neumann, and M. Qvenild, "Goodbye Intuition," Research Catalogue, 2020. [Online]. Available: <https://www.researchcatalogue.net/view/974962/974963>

INFORMATION EMBEDMENT IN LIVE-ELECTRONICS: SCORES, SETUPS AND PERFORMER KNOWLEDGE

Alex Hofmann

Department of Music Acoustics
mdw – University of Music and
Performing Arts Vienna
hofmann-alex@mdw.ac.at

Sonja Stojak

Department of Music Acoustics
mdw – University of Music and
Performing Arts Vienna
stojak@mdw.ac.at

ABSTRACT

Works for live-electronic music performance depend on a substantial amount of information about the music itself but also about the required technical setups. This paper examines materials for two live-electronic interpretations of historic tape music repertoire and four original works for live-electronics, created by different live-electronic partitioners (LEPs) within the artistic research project “Études for live-electronics”. Although written musical scores were not a requirement, for four of the works such scores were created. An analysis of the scores revealed different levels of information detail, depending on the relation of the LEP to the intended performance practice. Implications of access to the remaining information, often embedded in technical setups or retained as the performers’ background knowledge are discussed.

Keywords: Musical notation; Notation in electronic and electroacoustic music; Live-electronic music

1. INTRODUCTION

In music compositions until the 20th century, the score represented an important channel for information between the composers on the one hand and the conductors and performers on the other [1]. However, already with electroacoustic compositions, the function of a score changed. It became either a tool for the composers themselves to develop their music [2], or scores were written later, often by others, to analyse the works or to be used for pedagogical purposes [3].

With live-electronic music performances today, written music scores may or may not be created. In the former case there can be varying degrees of information about the music and the technical setups in this score, and in the latter case information may be contained in other domains such as in the live-electronic hardware and/or software setup including fixed media playback itself or in the background knowledge of the performer. This leads to the question of how creators of live-electronic works balance these domains when engaging in artistic research and when sharing their insights and materials.

Over the course of the artistic research project “Études for live-electronics” (2022–2026) at mdw – University of Music and Performing Arts Vienna, our aim was to gain a better understanding of the creation of live-electronic music by focusing on three relevant elements: a) setups and interfaces [4, 5], b) performance preparation procedures [6], and c) repertoire for live-electronic performers. This paper specifically focusses on the last element, although all three are often strongly linked, as will be discussed in Section 5.

2. METHOD

Starting from collaborative live-interpretations of historic electronic tape music compositions (Leiyla and the Poet (1959) by Halim El-Dabh; Stria (1977) by John Chowning; I of IV (1966) by Pauline Oliveros), we invited four external live-electronic practitioners (LEPs) to create new live-electronic artistic works. LEPs were introduced that it is foreseen for them to give a public 4-day workshop, in which they will share their artistic research approach, their performance preparation processes, and to give the participants the opportunity to acquire new repertoire. None of the LEPs got instructions to provide a written music score. The invited LEPs were a) Vietnamese multimedia composer and electroacoustic improviser LUÔNG Huệ Trinh b) Kenyan sound artist KMRU, c) Chilean visual artist, dancer and researcher Constanza Piña (aka Corazón de Robota), and d) Belgian-Congolese composer, musician and mastering engineer Cedrik Fermont (aka C-drik, Cdrk).

Although it was not required for the LEPs to work with musical scores, some decided to develop written scores, while others focussed on setups and interfaces in a ‘composed instrument’ fashion [7]. In order to explore the level of information detail within the delivered material, we describe and analyse the artistic works, including those derived from tape music repertoire (Section 3.1), as well as the new live-electronic works consisting of performance materials with a written musical score (Section 3.2) and without (Section 3.3).

3. ARTISTIC WORKS

This section provides an overview of the materials of the works. The complete materials of the works, including recordings, are available online [8].

3.1 Scores developed for live-electronic performance of historic tape music repertoire

3.1.1 *Leiyla and the Poet* (1959) by Halim El-Dabh

Leiyla and the Poet is a tape piece by Egyptian composer Halim El-Dabh from 1959 with a duration of 5 and a half minutes. During its premiere in 1961, El-Dabh put "elongated figures [...] with cloth on them" on stage to compensate for the absence of performers in electronic music [9], which inspired our artistic research team to work on a live re-performance of the piece with an actual live-electronic music ensemble.

A score was developed by the first author to coordinate the live performance parts, along with the extracted vocal stems from the original tape recording [10]. To develop the score, a transcription of the recording published in 1964 [11] was made with the iAnalyse Software [12, 13], including comments about El-Dabh's original material from the musicological analysis by Brigid Cohen [9]. Figure 1 shows the first page of the transcription score, with its signature beginning of the high-pitched sinusoidal sounding reed flute sample (green), followed by an entrance of processed vocal samples and instrumental recordings of a frame drum cadence, before at 0:54 the name "Leiyla" gets introduced by a register changing voice recording of El-Dabh himself. The score follows the principle of showing a symbolic transcription of relevant events, with additional comments from the Cohen analysis text. Performers are free to choose either live-electronic setups or other acoustic instruments to re-perform the piece.

During a 4-day workshop before the performance at the Kilele Summit 2024 in Nairobi (Kenya) with experienced LEPs (N = 3), prepared live-electronic patches in Pure Data were shared¹ that would allow to re-perform the piece. This included the hand-tracking setup based on *Cloud Hands* [5], that Nyokabi Kariuki performed during the concert performance to mimic the high-pitched reed flute sound. In the rehearsals prior to the performance, each performer decided for elements in the score that they would perform. For a quartett performance (Nyokabi Kariuki: Pure Data live-electronics, mbira; Labdi Ommes: orutu; Kostia Rapoport: Live (by Ableton) live-electronics; first author: tenor saxophone) of this piece in February 2024 at 'The Mist' in Nairobi (Kenya), the score was projected to the wall, visible to both the performers and the audience. Additionally, there were four public contributions to the wider audience during the Kilele Summit 2024 [14]: 1) a "Listening Session" (1h) on the music of Halim El-Dabh by Nyokabi Kariuki; 2) a 2h workshop session on "Building Live-Electronic Instruments" presented by Tim-Tarek Grund; 3) a 30 min talk on "Randomization Techniques in Live-Electronics Performance using Ableton Live" by Kostia Rapoport; as well as 4) a 2h workshop on "Body Percussion and Live-Electronics" by Nyokabi Kariuki and the first author.

¹ <https://github.com/grundton/Building-live-electronic-instruments-kilele>

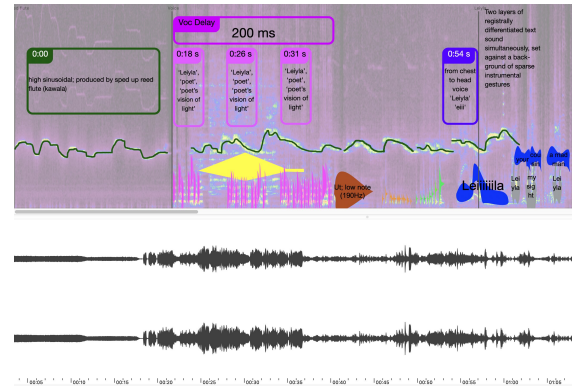


Figure 1. First page (1/5) of the transcription score for live performing Halim El-Dabh's *Leiyla and the Poet* with an ensemble using live-electronic- and acoustic instruments.

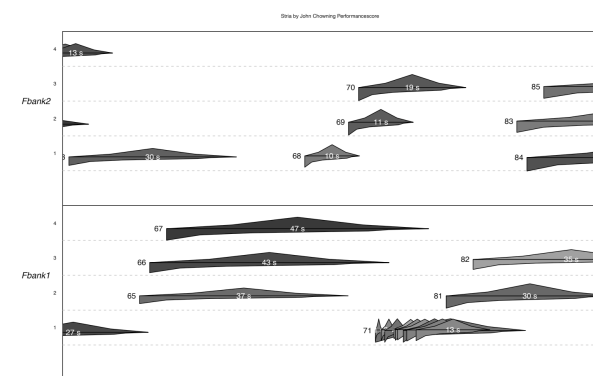


Figure 2. Excerpt (page 3/10) from Duo live-electronic performance score for *Stria*, showing the sound events with its duration and their assignment to the faders of two MIDI controllers. The brightness of the events indicates the fundamental pitch (bright = high frequency, dark = lower frequency) and the form of the events shows the respective modulation envelopes.

3.1.2 *Stria* (1977) by John Chowning

Stria is an algorithmic computer music composition by the American composer John Chowning that was premiered in 1977 as a quadrophonic tape (fixed media) performance with a duration of approx. 16 minutes. Based on the Csound reconstruction [15, 16] of the piece, a Max (by Cycling '74) and an open source Pure Data (PD) patch, that allow to live perform the piece as a Duo performance, in an event-after-event triggering approach using standard MIDI controllers, preferably with faders, was developed. To coordinate the actions of the two performers and to give an overview of who is triggering which event, a written score was derived by the research team at mdw from the Csound Score (a 31 * 383 data table with synthesis parameter settings) using the plot functions of the R-Statistics Software [17], resulting in a 10 page document (see Figure 2).

The created score shows the order of the sound events of the original composition together with the assignments to the faders on the MIDI controllers. Although it gives an indication of the length of each event, this score is omit-

ting an exact timeline at the bottom for interpretational freedom. The corresponding Max/PD patch runs the frequency modulation (FM) sound synthesis functions and has all synthesis settings stored in a table. These are read out, depending on the fader movements on the MIDI controllers by the performers.

In a 4-day workshop at mdw in October 2023 with students in electro-acoustic composition (N = 6), the prepared materials for the live performance version of *Stria* were provided with the aim for them to practise the piece. All students had prior experience in working with Max (by Cycling) and MIDI controllers. On the 4th day, they felt comfortable to perform *Stria* in a public concert.

3.2 Scores developed for new live-electronic performances

3.2.1 *How can I be tender?* (2023) by LƯƠNG Huệ Trinh

Vietnamese multimedia composer and electro-acoustic improviser LƯƠNG Huệ Trinh created a work called *How can I be tender?*. In order to develop the work, she collaborated with our project team on developing the digital musical instrument (DMI) Grain Bau, a granular effect sample player in Pure Data with a dedicated hardware interface, that borrows certain performance gestures from the traditional Vietnamese instrument Đàn Bầu [6]. The work was created to be performed by another artist than LƯƠNG herself, hence she delivered a visual descriptive score with action elements and a short textual instruction for the performer (see Figure 3). The score was made with the paid version of the online graphic design tool Canva (by Canva Pty Ltd.).

The work has a total duration of 11 minutes. The 3-page score, with 3 rows per page, gives the length dedicated to the playing duration of each notated row. The structure of the work can be recognized by the different colours and shapes of the graphic elements, which express different sound voices and sound characters. The colors and the shapes illustrate that the beginning and the end of the work have the same character, framing the work as a whole. Using the same colors with thinner shapes in the second row, marks a transition to a contrasting section of the composition. The next section (third row) is in contrast with the previous red and black colors through the use of pastel colors, softer shapes, and the addition of more musical elements, such as melody and rhythm. Although the action moments for playing certain samples are clearly indicated via small black boxes, there is no information or indication in the written score about the concrete sound material.

The sound material (12 audio samples with field recordings, melodic, and rhythmic sounds are provided with the software setup described in [6]) is not set as fixed materials of the work, so that each performer can create their own sample materials on a voluntary basis. The score outlines the structure, form and character of the work, as well as the information on durations, dynamics, and sample action moments within the given time frames. In order to perform the work, different performers would need time to create the samples that would be used (in case the provided sound material is not used), familiarize themselves with the

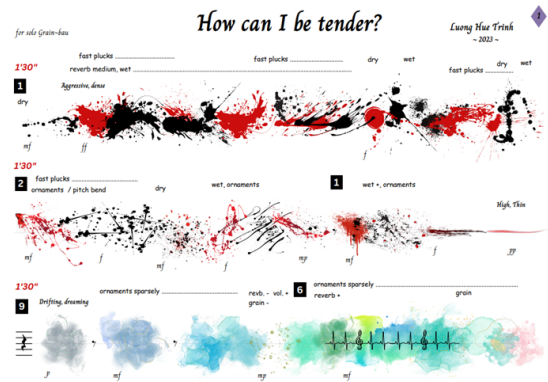


Figure 3. First page (1/3) of LƯƠNG Huệ Trinh’s work *How can I be tender?* for solo Grain Bau, showing the character of the music and its change from sharp tensions (black-red visuals) to the softer part (smooth pastel visuals).

granular effect sample player, eventually build and practise with the Grain Bau interface (alternatively use the standard MIDI keyboard controller option developed by [6]), or develop their custom interface. The work was premiered in 2023 by Tim-Tarek Grund and re-performed (using the provided sample material) in 2025 by the professional Đàn Bầu performer NGÔ Trà My (from the Vietnam National Academy of Music, Vietnam) at mdw.

3.2.2 *Beneath, Within* (2025) by Tim-Tarek Grund, Alex Hofmann, and Cedrik Fermont

The live-electronic trio performance *Beneath, Within* was created to explore the networks of human performers, cultural contexts, and AI, both in its production process and through its musical structure. The work was created by the artistic research project team for a collaboration with Cedrik Fermont and performed at The Kuala Lumpur Performing Arts Centre (klpac) in November 2025. For performance-related purposes, an action performance score was created in the Freeform Software (by Apple, Inc.), with the aim to serve as an orientation for the three performers and to indicate the time aspect, as well as providing instructions and specifying the points in time at which eye contact should occur to ensure coordination between the artists (see Figure 4) during the 12-minute performance.

In the score, there are no indications on the sound material that the performers have to use, since the saxophone and gong performers are mostly having improvised sections in different configurations with the AI-based RAVE (Realtime Audio Variational Autoencoder; implemented in Pure Data with nn~) live-electronics.

During the preparation process for this performance, musical patterns for the saxophone were algorithmically created that cover all pitches of the saxophone equally, so that the model is trained on the entire register of the instrument. These patterns are delivered separately in traditional music notation score, which was used during the recordings of the training data with the tenor saxophone. With these recordings (total duration approx. 3h) the RAVE tone trans-



Figure 4. The excerpt of the Trio performance score *Beneath, Within* coordinating the performer actions within the foreseen time frame, as well as the specific points for eye contact between the artists for a synchronised performance.

fer AI model was trained (training duration approx 5 days). To train the gong model, a less systematic approach was needed, as the variation in pitch is limited to the amount of gongs used.

As ethical use of AI in music composition is the conceptual key element of this work, it is encouraged that performers who have an interest in performing the work record and use their own data to train the models for their performances. In a public workshop in November 2025 at the Univeriti Malaya (Kuala Lumpur, Malaysia), the majority of participants (N=19) were students in music performance and sound engineering with prior experience in working with digital audio workstations, such as Cubase (by Steinberg) or Reaper (by Cockos). Participants were introduced to the concept of the performance and got insights on how to train and perform with RAVE tone transfer AI models using the required Open Source Tools (Pure Data; Python). Due to time constraints it was only possible for the workshop participants to experiment with the models provided by the workshop presenters, as there was not enough time for them to record training data and train their own models. However, with the tools provided, participants were enabled to train and work themselves with RAVE tone transfer AI models after the workshop.

3.3 Materials for live-electronic performances without a written musical score

3.3.1 *Ghost Tones Captured in a Khipu* (2024) by Constanza Piña

The artistic work *Ghost tones captured in a Khipu* was created by the Chilean visual artist, dancer and researcher Constanza Piña (aka Corazón de Robota). Together with the research team at mdw, the required live-electronic antenna setup was developed based on two things: 1) the design of the ‘Khipu’, a pre-Hispanic textile data storage system and 2) the electronic music piece *I of IV* (1966) by Pauline Oliveros. For the latter, the original electronic setup available to Oliveros at the University of Toronto Electronic Music Studio in July 1966 was emulated in Max (by

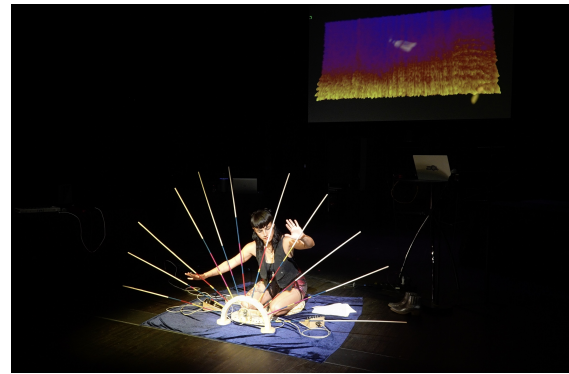


Figure 5. Constanza Piña performing *Ghost tones captured in a Khipu* at MAPPER Symposium at mdw Vienna in July 2025. The setup consists of her ‘Khipu’-inspired live-electronic antenna interface that controls a Max (by Cycling ’74) patch running an emulation of the ultra-sound oscillator setup used by Pauline Oliveros to create *I of IV* in 1966.

Cycling ’74), based on the description and schematics provided on the CD cover (see [18]).

In the work *Ghost tones captured in a Khipu*, the electro-textile interface acts as a physical bridge designed to channel, the essence of two fundamental texts: Pauline Oliveros’ *Quantum Listening Manifesto* [19] and Alan Poma’s *Andean Futurism Manifesto* [20]. Constanza Piña’s artistic approach is grounded in a technology-mediated invocation, in which words and sounds create a unified system. Her lecture-performance involved translating the hidden gestuality from Oliveros’ *I of IV* into a setup, capable of capturing all involuntary gestures within the electromagnetic field, making it possible to hear a ghostly order. Piña’s work executes the Manifesti sonically, transforming the electrical signals into an act of living memory and future vision. In her 30 minute performance-lecture (see Figure 5), a spectrogram visualization was projected to the wall, so that the audience can follow this visual demonstration of the relationship between the physical body of the performer and the technological object.

3.3.2 *FRAUGHT* (2024) by KMRU

Kenyan sound artist KMRU (Joseph Kamaru) was invited to create a live-electronics performance within the framework of artistic research. Following a conceptual approach, that combined the theory of “Dérive” by Guy Debord [21] with the process of fermentation, in his work *FRAUGHT*, KMRU experimented with audio processing, focussing on distortion, and investigated how sound transforms over time by letting it ferment without specific endpoints. Besides the conceptual text about the work, audio recordings of the work, and some photos and screenshots of audio equipment, neither a written music score nor other details about the live-electronics setup were provided. However, in a presentation during the Kilele Summit 2024 [14] and a lecture-performance during the MAPPER Symposium at mdw in 2025, KMRU gave insights about the setup in response to questions from the audience. Nonetheless, the requirement

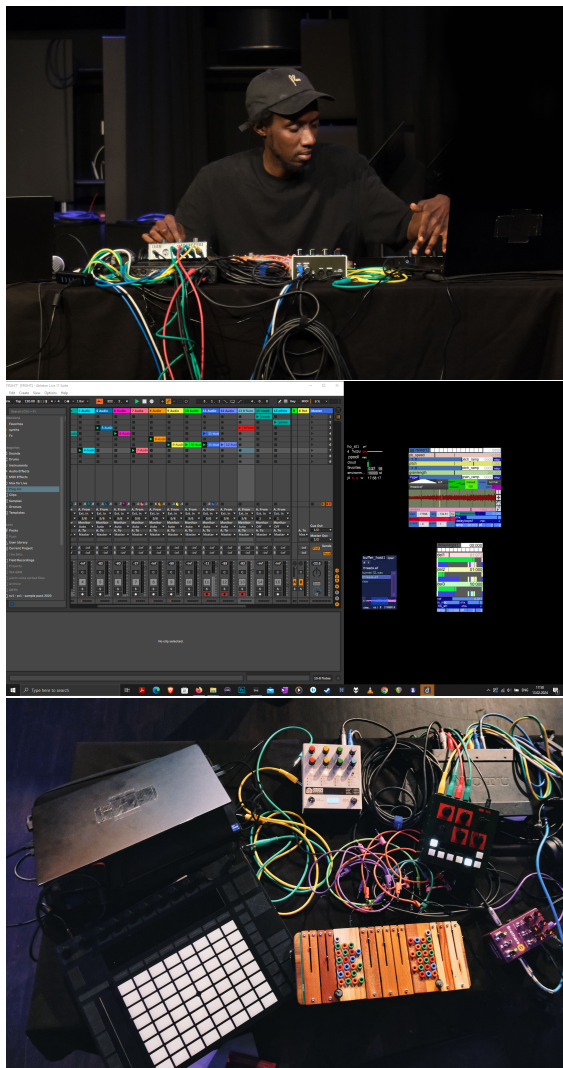


Figure 6. Top: KMRU performing *FRAUGHT* at MAPPER Symposium at mdw Vienna. Middle: Screenshot provided by KMRU of prepared live-set *FRAUGHT* in the Live Suite 11 Software by Ableton. Bottom: Live-electronic setup with (left) Computer running Live Suite 11 by Ableton and Push 2 Controller, (middle) Sidrax by Ciat-Lonbarde, Chroma Console by Hologram Electronics, (right) Audio Brothers AM pedal by Chase Bliss, 1010music Blackbox, and MOTU Audio interface.

for performing the work *FRAUGHT* strongly depends on the performer background knowledge and the proficiency with the live-electronic setup (see Figure 6) that he has.

4. ANALYSIS OF MATERIALS FOR THE WORKS

4.1 Function of scores in live-electronics

The four written scores (Section 3.1.1, 3.1.2, 3.2.1, 3.2.2) were each developed for works with a different number of performers, ranging from a solo live-electronics performer to a quartet with live-electronics and acoustic instrument performers. For the three works for ensemble performance (quartet – *Leiyla and the poet*, duo – *Stria*, trio – *Beneath, Within*) the main function of the developed scores was to

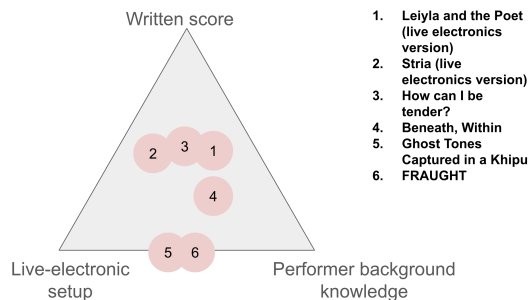


Figure 7. Location of information for the presented live-electronic works, distributed across the three relevant domains of the written score including written descriptions or explanations of the required technical setups, the live-electronic setup including fixed media such as samples and stems, and the background knowledge of the performer.

give time and action orientation for the performers to coordinate their actions.

Time and action instructions were also given in the score of the solo work *How can I be tender?* (Section 3.2.1). The function of the elements in this score aimed at providing more details about the structure of the work and the character of the sounds that a performer is intended to produce. This score shows a clear intention of aiming at different performers to learn the work with the score, while at the same time allowing a certain amount of interpretational freedom for the performers e.g. by choosing their own samples. In contrast, almost no interpretational freedom (except for the timing) for performers was given in the materials for *Stria*, especially by storing all relevant synthesis parameters in the live-electronic setup. A reason here is the concept of the piece written by Chowing, where form, pitches and sound spectra are algorithmically determined by strict mathematical rules derived from the Golden mean ratio [15]. Moreover, works where the creators also performed themselves, showed either abstract scores (Section 3.2.2) or no scores at all (Section 3.3.1, 3.3.2).

Therefore, the type of notation used, particularly the level of information detail, is influenced by the relationship between the creators of the work and the performance mode. This relationship depends on whether the creators perform the live-electronic works themselves or act in a more conventional composer-performer approach.

4.2 Information embedment in live-electronics

Considering that different levels of information detail for performers are presented in the written scores (or no scores were created) the information of the musical works must be embedded in other domains, taking into account that all works were already performed in live concerts. In live-electronics, relevant remaining domains are the setup (incl. fixed media) and the background knowledge of the performer(s) including their work-specific knowledge, experience in the interaction with the technical setup, and their experiences from previous performances of the work.

Figure 7, shows a map of the earlier discussed works, positioned by the amount of information stored in the respective domains: (top) the written score including written descriptions or explanations of technical setups, (left) the live-electronic setup including fixed-media and (right) the performer background knowledge.

For artistic works that were intended to be performed by other performers than its original creator, such as *How can I be tender?* (3), *Leiyla and the Poet* (1), and *Stria* (2), a written score was created containing varying levels of information, in comparison to works that were solo performed by the creators (*FRAUGHT* (6), *Ghost Tones captured in a Khipu* (5)), where all information were distributed between the live-electronic setup and the background knowledge of the performer. In the case of ensemble performances (*Leiyla and the Poet* (1), *Stria* (2), *Beneath, Within* (4)) the written score predominantly gives timing orientation for the performers. Thus, all the live-electronic works discussed in this paper fall into the middle and lower part of this map. This is because a significant amount of information is either contained in the live-electronic setups (which were not described in much detail in the scores, although some were shared as software online) or remains the background knowledge of the performers.

The degree of background knowledge sharing varied between the practitioners. However there seems to be an impact of eventual economical reasons. Some practitioners, who see more of their activities in live performance may be less open to share their materials than practitioners who see themselves more in the area of composition. Additionally, it seems that there may be regional effects, as practitioners who are registered with royalty collection societies receive fees for their works being performed by other performers. In contrast, practitioners who are not registered may wish to perform their own music themselves, as they depend on performance fees.

5. DISCUSSION

Our main observation, that even for works where written scores were created, a significant amount of information is either stored in the live-electronic setup or can only be provided by the performers themselves (in some cases also the creators of the works), is in line with earlier reports on the specifics of contemporary music created for electronic instruments [22]. The scores for live-electronics that were analysed in this paper, provided instructions for performing original works (Section 3.2) and to re-perform historic tape music compositions (Section 3.1). Here, LEPs used a variety of different software tools to create these scores, with some software being dedicated for music scoring (iAnalyse, by [12, 13]), and other software coming from other domains such as graphic design (Canva by Canva Pty Ltd.; Freeform Software by Apple, Inc.) or data analysis and visualisation (R-Statistics Software by [17]). All scores focused on giving temporal instructions for required performer actions, in most cases direct performer interactions with a dedicated live-electronic setup. These instructions were mostly on an abstract level, leaving much room for inter-

pretations of the works by performers or even improvisations of the performers. These observations suggest that live-electronic works are not fully specified by scores alone, but driven from the interaction between score, setup, and performer knowledge, with the configuration of the setup playing a central role in shaping this interaction.

For the case of our experiments on live performing historic tape music compositions, interpretation already played a significant role during the process of translating the works into live-electronic setups. This included the general choice of computer music languages [23], the specific signal processing implemented in these patches derived from preserved schematics, and the choice of possible performer configurations (solo, duo, quartet) including the choice or development of interfaces and mappings. These developed live-electronic setups were shared during public workshops with the aim to establish an active performance practice around these works.

Works without a score, where the live-electronic setup is the main constraint of what can or will be performed, are often understood as within the practice of ‘composed instruments’ [7], stemming from an observation by Schnell & Battier made on how IRCAM practices changed over time since the 1970ies until the early 2000s. Recently, our understanding of a live-electronic setup has become even more nuanced by examining what distinguishes an acoustic instrument from most setups today, here specifically a Max (by Cycling ’74) patch, where the latter is seen as a digital-and-analogue assemblage [24].

In contemporary live-electronic practices, not every assemblage is oriented towards a musical composition. Today’s widespread access to live-electronic music (software) tools gives more people access to electronic music making, leading to a broader range of approaches beyond academic contexts, extending those discussed by Schnell and Battier [7]. As also observed in this project, these approaches include practices from free improvisation, popular music, digital arts, transmedia art, and sound art. In such contexts artists often experiment with DIY electronics, field recordings, and the re-mixing of samples, resulting in assemblages that may share similarities with sound installations, even when performed live by the artists themselves. An approach to examining the creation of such live-electronic setups is through the perspective of ‘musical assemblation’, which does not seek to replace established concepts such as ‘composed instruments’ or ‘DMIs’, but rather complements them by considering assembly practices based on the configuration of existing audio software tools (e.g. audio plugins in Live (by Ableton) or Reaper (by Cockos)). Our concept of ‘musical assemblation’ is useful in this context because it highlights how, in many cases, live-electronic works may emerge from the configuration of tools, rather than from the design of a single instrument.

Across the works discussed in this paper, the proficient level of background knowledge of the performers about their assemblage remains a key source of information about how to perform their works. This information was often only available to participants during the workshops and performance-lectures, where they could interview the per-

formers directly. While this reflects established traditions of practice-based knowledge transfer in music education, it also highlights challenges for documentation and reproducibility. We therefore encourage performers to explore different forms of sharing their knowledge as additional materials with their works, such as explanatory texts, instructional videos, in public interviews, together with (open source) software including required samples, to externalise this knowledge. Such materials, independent from its format, may later be considered a form of score, if it meets the requirements of a) giving others sufficient detail for allowing to learn and re-perform the specific work without contacting the creator and b) being intended to exceed the creators lifetime. Possible ways to further explore here may be to store digital materials in repositories for long term preservation, providing detailed schematics of the assembled setups and patches etc., and by working with open source software that allows to study the inner workings of the components of an assemblage (UGens, objects, plugins, etc.) that determine the actual sounding result. We consider it essential to address these challenges for ensuring the long-term preservation and accessibility of live-electronic works.

Acknowledgments

We would like to thank all artists who contributed to this artistic research project for their creative work and for giving us insights into their working process. We are also thankful to the research team at mdw, especially Tim-Tarek Grund, Dustin Zorn, Karlheinz Essl, and Alexander Mayer for supporting the artists and the workshops. We are thankful to the participants, who took part in and contributed to the workshops and concerts. Furthermore, we would like to thank all project advisory board members, namely Martin Kaltenbrunner, Julia Mihály, Andrew McPherson, Marko Ciciliani, and Joachim Heintz for their support, especially in selecting the external LEPs. We are thankful to our international collaborators David Tinning from Santuri East Africa and Bernt I. Wærstad for their contributions at Kilele, as well as Marzelan Salleh, Hanafi Hussin, Roan Opiso, Ainolnaim Azizol, and Tazul Izan Tajuddin for advising us on questions regarding Southeast Asian instruments and for providing the opportunity to host a workshop and concert at their institutions.

This research was funded in part by the Austrian Science Fund (FWF) [10.55776/AR743] and in part by the ASEA-UNINET programme of the OeAD by the Austrian Federal Ministry of Women, Science and Research.

6. REFERENCES

- [1] H. Zemanek, “Aspekte der informationsverarbeitung und computeranwendung in der musik,” in *Experimentelle Musik*, F. Winckel, Ed. Schriftenreihe der Akademie der Künste, 1970, vol. 7, p. 68.
- [2] E. Minetti, “Exploring through writing: Operative moments in bruno maderna’s compositional process for “musica su due dimensioni” (1952).” Leiden, The Netherlands: Brill | Fink, 2023, pp. 209 – 230. [Online]. Available: <https://brill.com/view/book/9783846767146/BP000016.xml>
- [3] M. B. Corral, “The aural score as an educational strategy for an integrated pedagogy of arts,” *Innovative and Creative Education and Technology*, pp. 83–86, 2017.
- [4] T.-T. Grund, “Polyomino interface for pitch lattices,” in *Proceedings of the International Csound Conference (ICSC)*, 2024, pp. 169–170. [Online]. Available: <https://doi.org/10.21939/icsc2024>
- [5] T.-T. Grund and A. Hofmann, “Cloud hands: Three emanations of a touchless musical assemblage,” in *Proceedings of the 22nd Sound and Music Computing Conference (SMC2025)*, Graz, July 2025. Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Jul. 2025, pp. 108–114. [Online]. Available: <https://doi.org/10.5281/zenodo.15840925>
- [6] T.-T. Grund, T. H. Luong, and A. Hofmann, “Challenges and prospects in remote cross-cultural musical interface design,” in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2024, pp. 1–5. [Online]. Available: <https://doi.org/10.5281/zenodo.13904905>
- [7] N. Schnell and M. Battier, “Introducing composed instruments, technical and musicological implications,” in *Proceedings of the NIME*. National University of Singapore, 2002, pp. 1–5.
- [8] A. Hofmann, “Materials of ‘Études for live-electronics’ artistic research project,” <https://www.researchcatalogue.net/shared/35f74ec8577b3a47c46e154a715f0351>, 2025, accessed: 2025-12-19.
- [9] B. Cohen, “Sounds of the cold war acropolis: Halim el-dabh at the columbia-princeton electronic music center,” *Contemporary Music Review*, vol. 39, no. 6, pp. 684–707, 2020.
- [10] A. Défossez, N. Usunier, L. Bottou, and F. Bach, “Music source separation in the waveform domain,” *arXiv preprint arXiv:1911.13254*, 2019.
- [11] H. El-Dabh, “Leilya and the poet,” On Columbia-Princeton Electronic Music Center. Sony Music Entertainment (2014, originally released 1964 as Columbia Masterworks), 1964.
- [12] P. Couprie, “analyse 5: Le développement d’un logiciel pour la musicologie numérique,” in *Journées d’informatique musicale*, 2019.
- [13] —, “Developing software for digital musicology: Inside an experimental laboratory,” in *International Computer Music Conference*, 2023, pp. 72–78.

- [14] M. Guttridge-Hewitt, “Rewiring nairobi,” *DJ Mag*, no. 654, pp. 71–78, 2024.
- [15] K. Dahan, “Surface tensions: Dynamics of Stria,” *Computer Music Journal*, vol. 31, no. 3, pp. 65–74, 2007.
- [16] O. Baudouin, “A reconstruction of Stria,” *Computer Music Journal*, vol. 31, no. 3, pp. 75–81, 2007.
- [17] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: <https://www.R-project.org/>
- [18] P. Oliveros, “I of IV,” electronic works 1965 + 1966. PARADIGM DISCS 1997, 1966.
- [19] —, “Quantum listening: From practice to theory (to practice practice),” *Culture and Humanity in the New Millennium: The Future of Human Values*, pp. 27–41, 2002.
- [20] A. Poma, “Andean futurist manifesto,” *Lima, Peru: Soma Publications*, 2020.
- [21] G. Debord, “Theory of the dérive,” *Visual Culture: Spaces of visual culture*, vol. 3, pp. 77–82, 1958.
- [22] S. Berweck, “It worked yesterday: On (re-) performing electroacoustic music,” Ph.D. dissertation, University of Huddersfield, 2013.
- [23] A. McPherson and K. Tahiroğlu, “Idiomatic patterns and aesthetic influence in computer music languages,” *Organised Sound*, vol. 25, no. 1, pp. 53–63, 2020.
- [24] J. Snape and G. Born, “Max, music software and the mutual mediation of aesthetics and digital technologies,” in *Music and Digital Media: A planetary anthropology*, G. Born, Ed. UCL Press, 2022.

TOWARDS A NOTATION OF SCENO-ELECTROACOUSTIC WORKS

Rémy Jannelle

Université de Montréal
Laboratoire formes • ondes (LFO)
remy.jannelle@umontreal.ca

Nicolas Bernier

Université de Montréal
Laboratoire formes • ondes (LFO)
nicolas.bernier@umontreal.ca

ABSTRACT

This article addresses theatricalized musical performances within contemporary artistic practices. It examines how these works can expand beyond instrumental execution to incorporate staging and scenography as integral components of performance. The study proposes an analysis of the score of a *sceno-electroacoustic* work, a mode of musical practice in which sound emerges from transformations of the scenic space itself. This article contributes to reflections on notation, performance, and the role of scenography in electroacoustic music.

1. INTRODUCTION

From the second half of the twentieth century onward, contemporary musical performances have undergone a process of theatricalization [1, 2, 3, 4]. A few notable examples can be found in the work of Mauricio Kagel, *Match* (1964), *Zwei-Mann-Orchester* (1971–1973); John Cage, *Untitled Event* (1952), *9 Evenings: Theatre & Engineering* (1966); artists associated with the Fluxus movement, including Nam June Paik and his *Concerto for TV Cello and Videotapes* (1971); Karlheinz Stockhausen, *Gruppen* (1955), *Licht* (1977–2003); or, more recently, works by Alexander Schubert, *Terminal Infinity* (2025), *Eternal Dawn* (2025); Martin Messier, *1 Drop 1000 Years* (2023); Amy Chan, *Electromagnetic River* (2024); and Alexis Langevin-Tétrault, *Replika* (2025), *Interférences (String Network)* (2016).

The interpretation of such works extends far beyond the instrumental domain as staging and scenography are integral parts of the performance. This text proposes an analysis of the score of the *sceno-electroacoustic* work *Sémiographie*. Sceno-electroacoustic being a term introduced here to designate multimedia and interdisciplinary performances that foreground electroacoustic musical play as articulated through scenography.

Scenography is understood as an artistic practice that organizes space, gestures, movement, light, and video,

arranges objects and sets on stage, and transforms the stage into a sensorial space [5, 6, 7].

Sceno-electroacoustic performances can thus be conceived as works in which music is played through transformations of the scenic space, in much the same way that certain audiovisual performances use video or light to modulate sound. In this case, scenography becomes a vector of musical control.

2. PROBLEM STATEMENT

In current scientific literature, there are few examples of musical scores for sceno-electroacoustic works. The music world seems to have taken little advantage of developments in movement or dance notation, particularly the types of notations derived from the Labanotation, or ways of notating scenography like the work of stage director Robert Wilson. In sceno-electroacoustic performances, many artists or artist-researchers have techniques for remembering their performances, but the idea of a score rarely appears. The interdisciplinary nature of this type of performance raises issues of graphic representation, which makes notation complex and scores polymorphic.

3. SÉMIOGRAPHIE

3.1. A Sceno-electroacoustic Performance

The work under consideration in this text is entitled *Sémiographie*. This sceno-electroacoustic performance uses live electroacoustic performance techniques —e.g. real-time sampling and triggering—, while integrating several scenographic elements, such as objects, movements in space, and gestures as vectors for musical play. In this work, music is primarily generated by triggering samples, loops, and live sound recordings. These recordings are then processed by FX chains whose parameters change according to scenic actions made by the performer. To enable this type of control, an electronic device was developed based on principles derived from *Digital Music Instruments* (DMI), as theorized by Miranda & Wanderley (2006) and Marier & Piché (2009, 2017) [8, 9, 10]. Unlike the classic principle of DMIs, which consists of capturing the musician's gestures and transforming them into sounds via sensors,

The work consists of three main elements. The first is a position and gesture capture system, made with pressure sensors (FSR), a piezoelectric microphone, and switches. To collect positions in space, the pressure sensors are installed on the floor and are activated when the performer walks on them. These are used to trigger audio and/or light events during the performance. The piezoelectric sensor is used to track the performer's gestures. The switches scattered throughout the space trigger audio loops when activated by a performative action. Therefore, the music can only be developed through a precise interpretation of scenic displacement by the performer.

The third main element is a dynamic microphone that captures the sound of the paper, the performer's voice, and their breathing. It is also used to trigger samples and control FX chains with envelope following.

4. THE SCORE

A notation system for gestures, movements, and interactions with stage props was developed to create a score for the performance and archiving of the work. The symbols and text on the score serve primarily as markers that guide the performer's musical intentions, describe the sound material used, and provide a corpus of gestures and movements to be performed (Fig. 1). This method still leaves room for some freedom of performative actions. For instance, while the score indicates when and where the paper globally needs to be deployed, the precise gestures, modulations and intensities leading to this deployment are left to the performer's discretions. In the text *La partition en arts vivants : regard(s) sur l'utilisation du mot « partition » chez des artistes québécois-es*, it is stated that “the performer is no longer simply an executor but is integrated into the creative project.” The idea of the “well-crafted piece” and the “authoritative score” is rejected [11]. This quote sums up the intentions behind the score of *Sémiographie* that does not dictate, but rather accompanies, inspires, and guides

the performer without freezing the work of interpretation, allowing the performer to make the work its own.

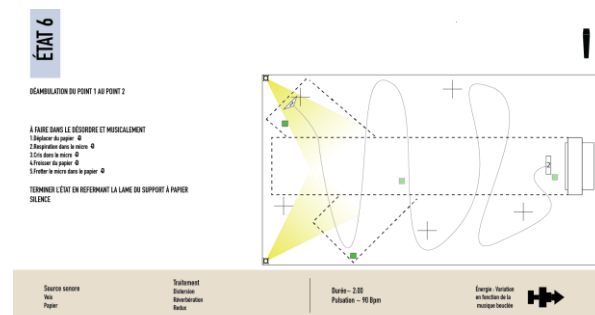


Figure 1. Excerpt from the score for *Sémiographie*.

3.2. Influences

In this sense, the score can be compared to Cléo Palacios-Quintin's notation work for her *Aléas* series. For example, for her piece *Aléas : Tableaux involontaires de la nature somnambule* (2020), the composer wrote a score combining the Western notation system, excerpts from poems by Thierry Dimanche, and drawings. When reading the score, there are indications that direct the musical action—at the top there are staves with notes—but a combination of expressions such as “éclats: mouvement vif” or “fleur jaune: délicat et aérien” placed under images of flowers and abstract shapes leave plenty of room for interpretation.

The score for *Sémiographie* is also in line with the work of artist-researcher Nicolas Bernier, who has been working for decades on the notation of performative works. The scores for his work *Transfert* 2017 (Fig. 2) combine text, symbols, and drawings indicating scenographic actions. On this score, Bernier drew the orientation of the lighting device that serves as a scenographic object during the performance. This way of noting the position of objects helped me develop this notation system.

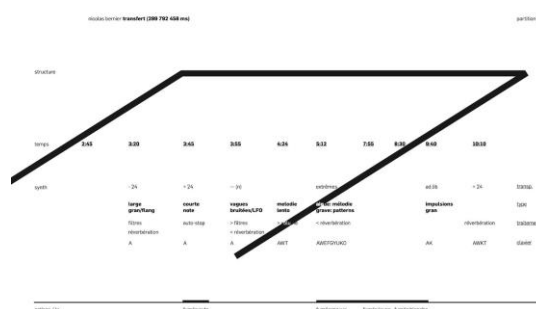


Figure 2. Excerpt from the score for Transfert (2017).
With the authorization of Nicolas Bernier.

4.2 Labanotation

The interdisciplinary nature of the work *Sémiographie* required the development of a type of notation that provides information on the music, scenography, lighting,

and movements. To achieve this, different forms of notation were used.

The form of writing that informed the notation of movement and positioning in space was Jacqueline Challet-Haas' *Symbolisation du mouvement* (Symbolization of Movement) [12]. This offers “a set of symbols drawn from Rudolf Laban's movement notation, facilitating observation, reflection, and experimentation with physical, spatial, and temporal factors.” [12] Challet-Haas's system has the particularity of “inducing movement proposals” [12], without dwelling on describing the gesture with the precision of Rudolf Laban's system. Although she uses signs from Labanotation, the dance notator proposes a system that leaves room for interpretation. It acts as a frame of reference for “describing the motivation behind a series of movements, their pattern [and] their structure” [12]. This feature makes it possible to write down movements and changes of position on stage while allowing the performer to improvise their gestures.

Challet-Haas uses lines and arrows to illustrate possible movements to be performed (Fig. 3). This way of representing movements in space has been taken up and adapted in the score for *Sémiographie* (Fig. 4). To indicate a suggested path on stage, thin lines are drawn. These are sometimes straight or curved and are used as a guide. The path may be different; the important thing is to move from one position to another.

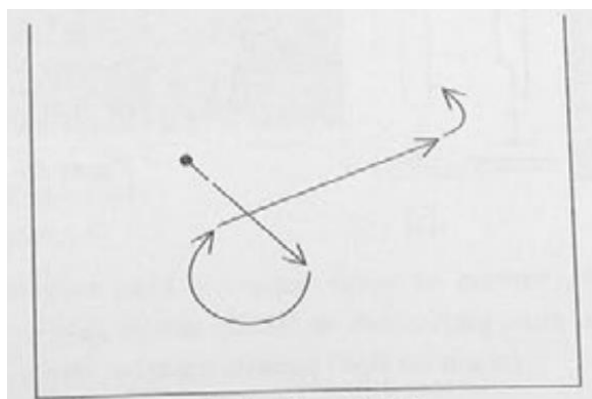


Figure 3. Excerpt from *La symbolisation du mouvement* issue de la *Cinétographie Laban* by J. Challet-Haas, p17.

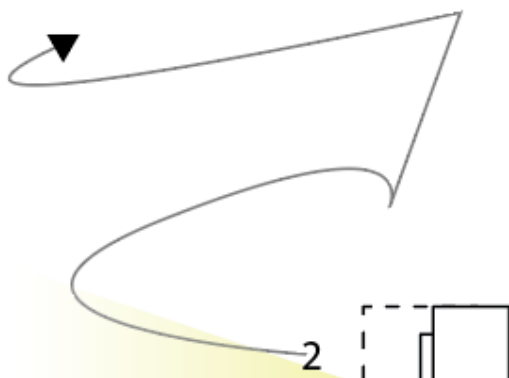


Figure 4. Excerpt from the score for *Sémiographie*.

Continuing with the idea of noting the interpreter's movements, Challet-Haas's method of noting the direction of travel is adopted and adapted (Fig. 5). Similar symbols are used to indicate the direction to be taken in space. For example, the stationary performer is represented by a rectangle (Fig. 1), while the moving performer is represented by a triangle (Fig. 1).

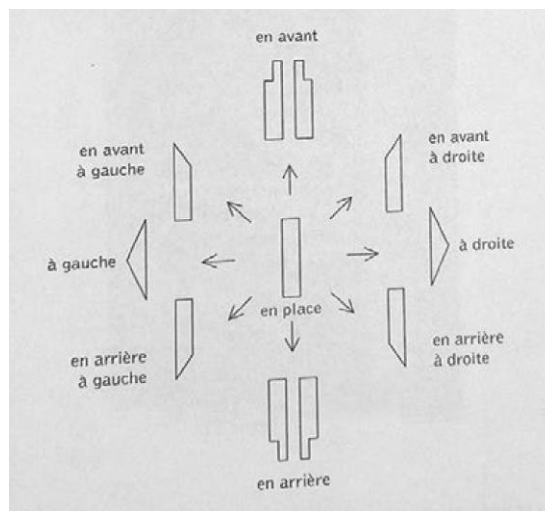


Figure 5. Excerpt from *La symbolisation du mouvement* issue de la *Cinétographie Laban* by J. Challet-Haas, p17.

Here, a distinction must be made between the two types of movement notation. Challet-Haas uses spatial reference symbols differently from those used in the score under study. The author uses an egocentric orientation. She indicates movements by referring to the position of the person dancing. In the case of *Sémiographie*, the movements are conceived in an allocentric manner. The position of objects on stage provides starting points and spatial references for movements. For performance purposes, the scenography is malleable; the stage can be reconfigured according to the space available during the performance. Movement indications are therefore relative to objects.

5. NOTATION STRATEGIES

5.1 States

The score of the work *Sémiographie* is divided into states (états), which are fragmented representations of the evolution of the stage space during the performance. These are important stages to be reached or moments of transition. This way of notating the sceno-electroacoustic performance allows it to be reproduced on a macroscopic and structural level, which in this context is a more faithful rendering than a microscopic representation. The interpretation is done through gestures, movements, and interaction with the device to interpolate between the different states. This way of structuring the performance creates a framework that allows for freedom of interpretation. The states serve as reference points, like anchor points that fix key moments, while what unfolds between them is more flexible and malleable. Each state

then contains stage-musical instructions. These instructions must be played in an improvised duration, but in an order that is indicated as follows:

Certain symbols will indicate certain types of musical actions. For example, a symbol representing a black microphone indicates that the microphone should be used by the performer, while a grayed-out symbol indicates that the microphone is not in use (Fig. 6).



Figure 6. Excerpt from the score for *Sémiographie*

There are also text-based instructions that provide context for the state while presenting a set of gestures to be performed. Figure 7 indicates that this is a state of wandering between points 1 and 2. The performer then moves across the stage through various scenographic objects. On the line below (Fig. 7), a clarification is made about the gestures. They must be performed “in a disorderly and musical manner.” This deliberately vague instruction serves to encourage a musical approach on stage. The small circular arrow accompanying each instruction suggests that the gesture can be repeated several times—the exact number being left to the performer's discretion—before moving on to the next one. The use of the term “musically” allows the performer to determine the order, intensity, and frequency of repetition of the gestures.

DÉAMBULATION DU POINT 1 AU POINT 2

À FAIRE DANS LE DÉSORDRE ET MUSICALEMENT

1. Déplacer du papier 🌀
2. Respiration dans le micro 🌀
3. Cris dans le micro 🌀
4. Froisser du papier 🌀
5. Frotter le micro dans le papier 🌀

TERMINER L'ÉTAT EN REFERMANT LA LAME DU SUPPORT À PAPIER SILENCE

Figure 7. Excerpt from the score for *Sémiographie*

The center right shows the stage layout as seen from above (Fig. 8). This section is the core of the score's scenographic representation. It shows where objects are placed, where the light comes from, where the performer should stand, where they should move, and their position in space. This representation contains several pieces of information about the stage layout. On the right, two superimposed rectangles represent the roll of paper. The four green squares correspond to the FSR sensors placed

on the floor: when they are pale, they are deactivated; when they are dark, they are activated. The long-dotted rectangles indicate where the paper should be. The five crosses mark the location of the switches. The Fig. 7 also shows a triangle with a “1” and a rectangle with a “2.” The triangle indicates the starting point of a movement, while the rectangle indicates the final position. Finally, the wavy line that crosses the entire representation shows that this movement can take place throughout the scene.

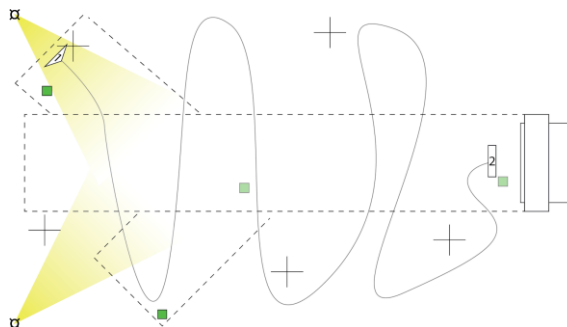


Figure 8. Excerpt from the score for *Sémiographie*

At the bottom of the score, on the horizontal beige band, information about the music is provided. On the left-hand side, you will find the type of sound source and the different treatments used during the state. On the right-hand side, an approximation of the state's duration in minutes and seconds is provided, along with the musical tempo, the energy form, and the musical intention. (Fig. 9). Each state has a symbol and a short phrase that suggests a musical energetic development. In Figure 9, several black lines are layered on top of one another to form a sideways triangle. This triangular shape suggests the idea of a crescendo, while the black lines imply a layering of sound textures. Other symbols are also used, for example a rectangle indicating an energetically stable state, or a triangle pointing to the right signifying a decrescendo.

Source sonore Voix Éclairage Synthèse	Traitement Brouillon Intérieur Exterieur	Durée - 2:00 Publication - 90 Sym	Énergie - Niveau et suppression de craques sonores, l'ensemble suggère
--	---	--------------------------------------	---

Figure 9. Bottom part from the score for *Sémiographie* which indicates global information of a state

6. CONCLUSION

This multimodal form of notation is a first attempt in my research-creation practice for writing scores for sceno-electroacoustic works. It allows me to record the articulation between movement, gestures, scenography, and sound, which is fundamental when the performance is no longer based solely on sound production.

However, the notation technique developed has its limitations, particularly in terms of its ability to provide precise information about artistic intention in different states. This raises questions about the issues of representativeness for this type of work.

As it is often the case in modern dance and theatre, video recording proved to be an effective way of filling these gaps. Where notation provides information on the structural aspects of the performance, video shows what the score can only suggest: the musical intention, the emotion conveyed, the way the body moves, or its posture. Video would thus complement the score, allowing the performance of the work to be reconstructed with greater fidelity.

In the context of the work *Sémiographie*, there is no video accompanying the score, to leave more room for improvisation and interpretation embodied by the performer (knowing that the work is meant to be performed solely by the composer himself). It is not important to have precise instructions; the performer is free to embody the performance as they wish.

The proposal developed here shall contribute to nourish the diversity of performative notation, while not conceived as an attempt towards standardized writing, since the diversity and specificity of the works make it very difficult for a common language to emerge. There is not yet a truly standardized performative notation system, since practices oscillate between graphic, textual, video, and even tablature forms [13]. It seems more relevant to imagine a link between shared principles and the specific characteristics of each creation, where different modes of notation complement each other to convey the information relevant to the reinterpretation of these works.

7. ACKNOWLEDGEMENT

This article is part of the Performative Notation project directed by Nicolas Bernier and funded by Fonds de recherche du Québec – Société et culture (FRQSC), a research project held at Laboratoire formes • ondes (LFO) at Université de Montréal.

8. REFERENCES

- [1] E. Salzman and T. Desi, *The new music theater seeing the voice, hearing the body*. Oxford University Press, 2008.
- [2] C. Salter, *Entangled: Technology and the Transformation of Performance*. MIT Press, 2010.
- [3] S. Dixon, *Digital Performance : A History of New Media in Theater, Dance, Performance Art, and Installation*. MIT Press, 2007.
- [4] N. O'Dwyer, *Digital Scenography : 30 Years of Experimentation and Innovation in Performance and Interactive Media*. Methuen Drama, 2021.
- [5] C. Guinebault, "Scénographie et représentation : une certaine façon d'appréhender le monde", *Études théâtrales*, vol. 2, no. 54-55, pp. 291-297, 2012.
- [6] M. Freydefont, S. Muston, and E. Gangloff. *Scénographier l'art*. Réseau Canopé, 2015.

- [7] C. Ouvrard, "La scénographie, un art de l'Incomplet", *Études théâtrales*, vol. 2, no. 54-55, pp. 110-112, 2012.
- [8] E. R. Miranda and M. M. Wanderley, *New Digital Musical Instruments : Control and Interaction Beyond the Keyboard*. A-R Editions, 2006.
- [9] M. Marier and J. Piché, "Musiques pour éponge : la composition pour un nouvel instrument de musique numérique", D. Mus. thesis, Université de Montréal, 2017.
- [10] M. Marier, "Le rapport geste-son en musique électroacoustique", M. Mus. dissertation, Université de Montréal, Montréal, Canada, 2009.
- [11] E. Giacomazzi, I. Héroux, and J. Bienaise, "La partition en arts vivants : regard(s) sur l'utilisation du mot « partition » chez des artistes québécois·es", *Percées*, no. 6, pp. 1-17, 2021.
- [12] J. Challet-Haas, *La symbolisation du mouvement issue de la Cinétopographie Laban*. Centre national de la danse, 2015.
- [13] J. Sermon and Y. Chapuis, *Partition(s) : objet et concept des pratiques scéniques (20e et 21e siècles)*. Les Presses du réel, 2016.

ITERATIVE SCORE FOR DIGITAL SOUND COMPOSITION

Peter Plessas

Department of Composition Studies and Music Production
University of Music and Performing Arts Vienna
plessas@mdw.ac.at

ABSTRACT

The composition of electroacoustic music for the recorded medium has become largely independent from the requirement to furnish a realization score. While digital technology invites intuitive compositional work in graphical audio software, a documentation of such creative activity often requires an a posteriori analysis and reverse-engineering approach. The iterative score presented here attempts to bridge the prescriptive and the descriptive nature of scores for sound compositions, that is musical works based on the transformation of recorded sound. Instructions for repeated realizations of such a composition are given in a case study, embracing a subtractive approach to problem solving with technology and providing a resource with good accessibility for visually impaired musicians.

1. INTRODUCTION

Textual or graphical scores for electroacoustic music on the recorded medium are frequently differentiated between prescriptive and descriptive scores. Prescriptive scores give partial or full instructions to realize a given work. They may also serve as a kind of parallel notation for the composer herself, used as a reference in production in order to decouple some of the decision making from the realization, while keeping track of the overall structure of the work. Descriptive scores for analyses and informed listening may employ visual representations [1] and inventories and supplementing texts [2]. They are surprisingly often created by individuals other than the composer or as a team effort. Automated transcription techniques for electroacoustic music on the recorded medium do frequently produce visual representations modeled after human auditory perception [3]. Such representations are useful in music analysis and as pedagogical support when followed along while listening as many visual scores feature a time axis similar to traditional symbolic scores. The informative value of time-varying spectral information may be questioned for its inability to group partials into events and to organize such events into auditory streams as described for the human auditory system [4]. Visual information from spectral analysis can successfully complement listening by showing in greater

detail what may not have been perceived or identified by ear.

2. REALIZATION SCORES IN ELECTROACOUSTIC MUSIC

Prescriptive scores for the realization of electroacoustic music follow a wide variety of notation strategies. Early compositions submitted as application for an realization in the Cologne WDR electronic music studio were initially expected to adhere to a scheme for graphical notation [5] in order to be realized by the studio staff, not necessarily including the composer. The correlation of the audible result with the original written instructions may have been loose in such cases but these scores allow repeated realizations by other individuals, which is an easily overlooked feature. To what extent a realization by different individuals includes a subjective and artistic interpretation is an interesting question and open to further research. Its examination would certainly require multiple realizations of a work by various artists, and a differentiation of which variations are the result of unintentional influences of the technology, and which are the result of choice and taste.

John Cage's "Williams Mix" (1951-53) constitutes an early realization score for an octophonic tape composition using recorded sound. The indicated edits of the magnetic tape are influenced by random procedures and do not trace the composers stepwise decision making path. The score does not prescribe a single specific source material and hence does not allow the exact recreation of the composer's initial version, but invites individuals to realize new renditions of the piece with original sound material [6]. Automated implementations for *Williams Mix* have been created by others [7][8] allowing an easy exchange of the source sound material subjected to digital edits.

Several of Trevor Wishart's compositions can serve as examples using recorded sound as sole material. They have been described by the composer in written texts, concerning amongst other topics the sources, transformations and layout of the sound material [9][10][11]. A digital resource in [12], covering his work "Imago" (2002), includes references to the implementation of the sound transformations and to the assembly of the sounds into larger structures while also providing the actual audio files for study.

Audible influences of technological shortcomings in multiple renditions of a compositions can be reduced by working entirely in the digital medium. The move to such a workflow however reduces the need, or desire, for a realization by others, as unlimited cycles of design-rendering-evaluation of

digital music become possible. This workflow introduces the danger of getting lost in a vast array of possibilities and of being able to revoke all decisions by infinite “undo” steps in software [13]. Cheap and accessible personal technology furthermore helps the composer to carry out all production steps herself, eventually becoming a monolithic luthier-composer-performer, a *hyphenated musician* [14].

Realization scores in form of pure software code are often targeted at works predominantly realized with sound synthesis.¹ This owes to the fact that digital synthesis may be sufficiently well described in textual or graphical form using a limited and standardized set of digital signal processing (DSP) blocks. Digital synthesis scores are self-sufficient without any additional audio data provided [15], but do often exhibit a structure that does not reflect the chronological development, for example a division between synthesis instruments and score statements. When well documented with comments and explanations, such software code may nevertheless allow a purposeful study by a human reader. Software that permits the performance of music by live coding may furthermore permit a chronological tracking and playback of all text input by means of a history file [16]. Modifications to *composed instruments* [17] can likewise be analyzed as a kind of interpretation by their players, either in advance or during a musical performance [18].

For sound compositions in the tradition of acousmatic music, which use audio recordings as their primary source material, a realization score is generally difficult to furnish. This is especially—and paradoxically—the case when a musical work is created with a digital audio workstation (DAW). Graphical multi track editing and mixing software offers an intuitive way of working with many layers of recorded sound, permitting its precise arrangement in time and transformation with effect plug-ins. While such real-time plug-ins are the de facto way of shaping audio signals in a DAW, they are limited by not being able to access the entire length of the signal in one step due to their block-wise processing. This is a drawback to composers wanting to redesign the time axis of an audio recording with signal processing algorithms. The Audio Random Access (ARA) extension [19] to existing audio software standards permits such access, although plug-ins returning audio data of different durations to the host DAW may produce unforeseen conflicts in the time line. Despite the obvious advantages of real-time operation, composition has been outlined as being essentially a non-real-time process in literature [20].

3. THE ITERATIVE SCORE

An iterative score formalizes a stepwise realization of an electroacoustic composition using recorded sound. When the actual audio source material has been made available, the composition can be studied, re-created and optionally modified by individuals other than the composer. An iterative score will hence serve as a prescriptive as well as an descriptive score. Its descriptive value can be greatly enhanced when its stepwise instructions are furnished with text comments by the composer. The creation of an iterative

¹ Some of the earliest examples of digital music may have been distributed as code listings on printed paper

score for sound compositions is greatly facilitated by using software that offers the transformation of audio files with a text-based command-line interface working in non-real-time. Example software include the Composer’s Desktop Project (CDP) [21], Sound Exchange (SoX) [22] and to some extent scriptable graphical audio editors and any DSP language offering stepwise execution. Restricting oneself to software that uses the operating system command line interface (CLI) offers a standardized and modular interface to multiple sound transformation programs as well as to powerful file system operations. The most important CLI feature in this context is the built-in command history and its ability to export a list of all typed commands into a text file, which can then be redacted to yield the steps of an iterative score. These steps will then allow repeatable realizations, different or identical to the original work. An iterative score represents the minimum number of steps required to arrive at the final composition starting from the source audio material. It does to some extent give a chronological account of the creative path the composer has taken but does not give evidence of all the decisions a composer has taken of which some were eventually rejected. For example, *shadow edits* [23], which did not make it into the finished piece, but which were possibly necessary to arrive at the accepted edits, are undocumented. An iterative score should therefore not be considered a close representation of the composer’s way of working, but a sequence of steps that were accepted as suitable for a specific composition. Iterative scores however permit other individuals to inspect the intermediate results of the sound transformations and to alter technical parameters, observing their impact on the final sound material.

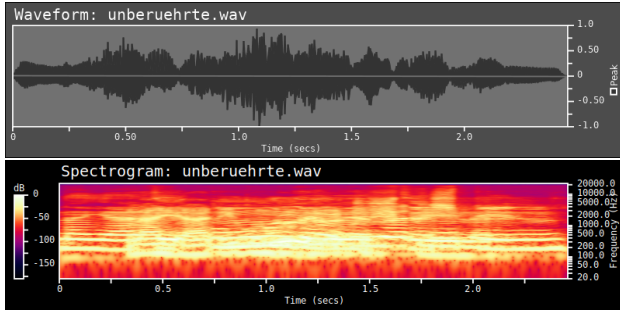
4. CASE STUDY

The feasibility of an iterative score that encompasses prescriptive as well as descriptive functions is demonstrated with regard to the author’s sound composition “Affair: Seven plus or minus two” (12 minutes, stereo) for the recorded medium. Commissioned for the dance performance “Affair” by Christine Gaigg in 2019, its title is inspired by George A. Miller’s publication regarding human capacity for processing information [24]. All five parts of this composition were realized almost exclusively using the command-line version of the CDP 7.1 software and short recordings of human speech taken from the performance text. CDP was chosen for its bewildering array of sound transformations targeted at electroacoustic music such as spectral or waveset processing, and for its GPL license and cross-platform availability. No operating system commands were used in order to keep the realization independent of the actual platform used. For the fifth part of the composition, “5. Code Adam”, the iterative score is presented next.

4.1 Source material

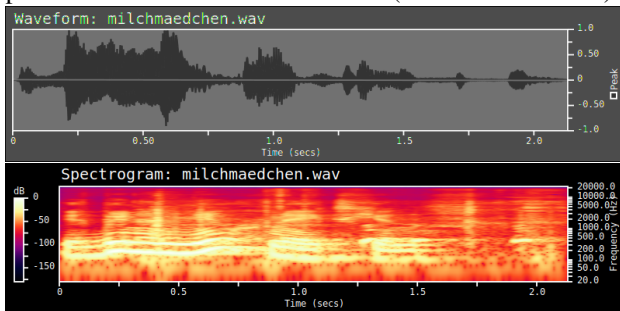
The source material is the mono file `unberuehrte.wav` (2.46 seconds), in which several male and female voices speak the german words “... die Unberührte spielen.” (“... playing the pristine.”) together. This recording has been se-

lected for its semantic content, the presence of unvoiced and voiced material and the somewhat common speech melody of the speakers.



Listen to [unberuehrte.mp3](#) online.

A second source file is `milchmaedchen.wav` (2.13 seconds) in which a male voice is speaking the german phrase “Ich bin ein Milchmädchen” (“I am a milkmaid.”).



Listen to [milchmaedchen.mp3](#) online.

4.2 Transformation steps

The audio file `unberuehrte.wav` is time-stretched and has its partials sustained by accumulation before it is subjected to a bank of time-varying resonant filters. The result is then chopped up into quasi-periodic chunks are each reversed in time.

The second audio file `milchmaedchen.wav` is changed to half-speed and a noisy, percussive and initially unintended trailing sound is extracted and reverberated.

Each of these transformations is iteratively described next, with hyperlinks pointing to the respective mp3 versions of the resulting audio files.

4.2.1 Spectral time stretch and accumulation

The audio file is converted to a spectral domain file (.ana) using the default window of 1024 samples and overlap 3:

```
pvoc anal 1 unberuehrte.wav unberuehrte.ana
```

Spectral time stretching with a factor 30 is applied, resulting in a file lasting 1 minute 14 seconds:

```
stretch time 1 unberuehrte.ana
unberuehrte-stretch30.ana 30
```

Listen to [unberuehrte-stretch30.mp3](#) online.

Energy in each frequency bin is sustained by *spectral accumulation*

```
focus accu
unberuehrte-stretch30-gain.ana
```

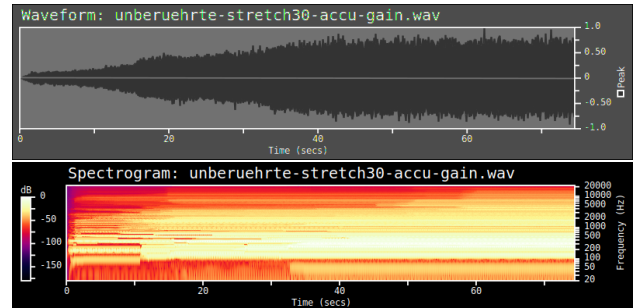
```
unberuehrte-stretch30-accu.ana
```

and converted back to the time domain:

```
pvoc synth unberuehrte-stretch30-accu.ana
unberuehrte-stretch30-accu.wav
```

This conversion will report a necessary gain factor of 0.388423 to prevent signal clipping, which is applied to the spectral file before the conversion to time domain:

```
spec gain unberuehrte-stretch30.ana
unberuehrte-stretch30-gain.ana 0.388423
pvoc synth
unberuehrte-stretch30-accu-gain.ana
unberuehrte-stretch30-accu-gain.wav
```



Listen to [unberuehrte-stretch30-accu-gain.mp3](#) online.

4.2.2 Time-varying resonant bandpass filters

A bank of time-varying resonant bandpass filters is overlaid, first querying the sound file duration

```
sfprops unberuehrte-stretch30-gain-accuN.wav
```

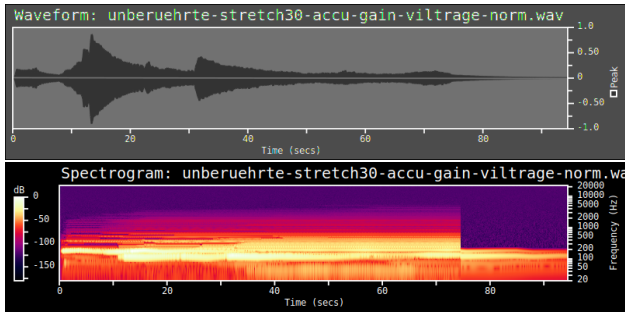
and adding 20 seconds allowing for a sufficiently long filter resonance, resulting in a new duration of 94.4461 seconds. The text file `viltrage.txt` is generated automatically, holding four column pairs of time and frequency values over the 94.4461 seconds for 4 resonant filters having pitches from 28 to 63 linearly distributed 1, randomized 0.5, minimum amplitude 0.5, amplitudes distributed decreasing with pitch -1 every 4 seconds and times randomized 0.5. A random seed -s123 is specified² to allow for identical random values on repeated runs of the command:

```
filtrage filtrage 2 viltrage.txt
94.4461 4 28 63 1 0.5 0.5 0 -1 4 0.5 -s123
```

A bank of resonant bandpass filters controlled by `viltrage.txt` with a very steep and resonating Q of 3000, a large make-up gain factor 500 and added tail of -t20 seconds is applied, followed by gain normalization:

```
filter varibank 2
unberuehrte-stretch30-accu-gain.wav
unberuehrte-stretch30-accu-gain-viltrage.wav
viltrage.txt 3000 500 -t20
modify loudness 3
unberuehrte-stretch30 [...] viltrage.wav
unberuehrte-stretch30 [...] viltrage-norm.wav
```

² The `viltrage.txt` used for the published version of this composition was created without a seed value and can hence not be recreated here.



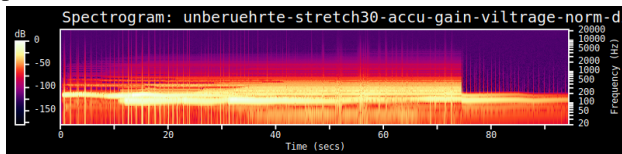
Listen to [unberuehrte-stretch30-accu-gain-vilrage-norm.mp3](#) online.

4.2.3 Reversal of wavesets

Every 30 wavesets are reversed as a group, with a waveset being the time section of the signal between three successive zero crossings [25], similar but not necessarily identical to a wavecycle:

```
distort reverse
unberuehrte -[...] vilrage -norm.wav
unberuehrte -[...] vilrage -norm-distRvs30.wav
30
```

The amplitude waveform for the file created does not show any significant change but its spectrogram does. It indicates the onsets of the reversed groups of wavesets, which are in fact discontinuities or impulses with a short broadband spectrum.



Listen to [unberuehrte-stretch30-accu-gain-vilrage-norm-distRvs30.mp3](#) online.

Variations of this transformation are produced, re-using the above command with slower pulsations of 50, 100, 125 and 300 wavesets, following the same output file naming scheme.

4.2.4 Slowing down playback speed and cutting

The file `milchmaedchen.wav` is slowed down to half speed and hence 4.26s duration:

```
modify speed 1 milchmaedchen.wav
milchmaedchen-speed05.wav 0.5
```

Listen to [milchmaedchen-speed05.mp3](#) online.

A soft percussive trailing sound is extracted by cutting out a section from 3.104 to 3.788 seconds, followed by gain normalization:

```
sfedit cut 1 milchmaedchen-speed05.wav
milchmaedchen-speed05-cut.wav 3.104 3.788
modify loudness 3
milchmaedchen-speed05-cut.wav
milchmaedchen-speed05-cut-norm.wav
```

Listen to [milchmaedchen-speed05-cut-norm.mp3](#) online.

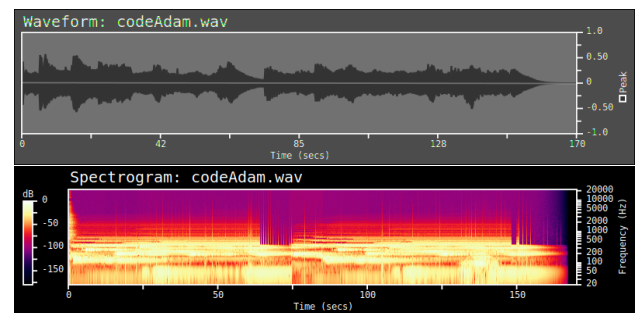
4.2.5 Assemblage

All files are renamed for better readability, showing only the last transformation affix. The files are assembled in a three track timeline observing the values from the edit decision list in Figure 4.2.5. Following the Position in the timeline, the Start and Length values therein indicate the time portion that is used out of that file. All values are in seconds, while Gain is in dBFS.

Track	File	Position	Start	Length
1	<code>milchmaedchen-cut.wav</code>	0	0	0.684
2	<code>unberuehrte-distRvs200.wav</code>	0.161	10.713	11.81
2	<code>unberuehrte-distRvs125.wav</code>	11.961	22.526	12.68
2	<code>unberuehrte-distRvs100.wav</code>	22.53	33.033	11.48
2	<code>unberuehrte-distRvs50.wav</code>	32.13	42.642	24.14
2	<code>unberuehrte-distRvs30.wav</code>	55.592	66.144	19.17
2	<code>unberuehrte-distRvs125.wav</code>	73.946	0	11.10
2	<code>unberuehrte-distRvs100.wav</code>	84.955	11.097	15.64
2	<code>unberuehrte-distRvs50.wav</code>	99.035	25.52	44.94
2	<code>unberuehrte-distRvs50.wav</code>	143.959	70.444	6.58
2	<code>unberuehrte-distRvs125.wav</code>	149.59	73.303	17.50
3	<code>unberuehrte-distRvs50.wav</code>	129.055	55.54	6.193
3	<code>unberuehrte-distRvs50.wav</code>	132.6	59.085	10.04
3	<code>unberuehrte-distRvs50.wav</code>	140.074	66.559	4.788

Edit decision list used in the assemblage.

Artificial reverberation with a long reverb tail is added to track 1 holding `milchmaedchen-cut.wav`. Cosmetic EQ is applied to all `unberuehrte...` files on track 2, reducing resonances at 122, 200 and 250Hz and slightly amplifying high frequency content in a wide band centered at 846Hz. The three `unberuehrte...` files on track 3 shifted by -24 semitones in pitch to create a low-frequency layer.



Listen to the published version of [codeAdam.mp3](#) online.

5. DISCUSSION

Several challenges do exist in the creation of iterative scores. All random procedures have to allow the specification of numerical seed values for repeatable results. Although the differences of unseeded random procedures might not be immediately audible, they might nevertheless cause a radically different result when subjected to subsequent sound transformations. The assemblage of sound files on multiple tracks in a time line can be comfortably done inside a DAW but has to be exported as an edit decision list containing only a subset of information for the iterative score. Although CDP offers powerful text-based mixing functions, they were not employed here due to their rather complex

usage.

An iterative score will produce many advantages, some of them unexpected. Its documentary approach allows the composer to re-use proven sound transformation techniques in other pieces and access a summarized work diary including sound examples. As such, the iterative score outperforms the undo history of the digital audio workstation, which contains all actions, but not necessarily all decisions taken.³ The score presented here furthermore poses an *open source score*, guaranteeing a certain longevity of a work as well as independence from commercial software. When published, iterative scores are of great value in analysis and education. For such a knowledge transfer to work, it is required that the artist lays open part of her way of working, which might not appeal to every composer. The example of an iterative score in this case study documents a rather straightforward way of composing electroacoustic music, with an advance creation of sound material by transforming digital audio files, a selection from these results and their arrangement on a time line on multiple audio tracks. Such a way of working is by no means representative of the plethora of strategies that are followed consciously, or unconsciously, by composers. Other strategies might feature a repeated back and forth between the generation of material and its selection and arrangement in a more circular composition path. Furthermore, time-varying parameters are not featured in the above example. Although technically possible with breakpoint text files in CDP, they result in a large amount of text data, especially when automation curves are entered with a haptic fader interface.

6. CONCLUSIONS

The approach towards a prescriptive as well as descriptive score presented here draws from a subtractive notion towards problem solving using technology [26]. Discarding the graphical user interface and computer mouse will permit the edition of a human-readable iterative score. Such a score constitutes a meaningful textual documentation, especially when provided with additional comments by the composer, and features enhanced accessibility for visually impaired individuals. The iterative score, as a simple text document, is robust to changes in digital file formats, to patent issues and the availability of commercial software. It furthermore opens up new possibilities for an analysis and re-interpretation of digital sound compositions by other artists.

7. REFERENCES

- [1] R. Wehinger, *Ligeti Artikulation. Hörpartitur*. B. Schott's Söhne, 1970, Reprint.
- [2] J.-C. Thomas, P. Mion, and J.-J. Nattiez, *L'envers d'une œuvre. De Natura Sonorum de Bernard Parmegiani*. INA-GRM, Buchet/Chastel, 1982.
- [3] P. Couprie, "Methods and tools for transcribing electroacoustic music," *Technologies for Music Notation and Representation (TENOR)*, pp. 7–16, 2018.
- [4] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.
- [5] H. Eimert, F. Enkel, and K. Stockhausen, "Fragen der Notation Elektronischer Musik," *Technische Hausmitteilungen des Nordwestdeutschen Rundfunks*, vol. 6, no. 1/2, pp. 52–54, 1954.
- [6] R. Kostelanetz, *Conversing with Cage*, 2nd ed. Routledge, 1987, 2003.
- [7] T. Erbe, "The Computer Realization of John Cage's *Williams Mix*," *Proceedings of the International Computer Music Conference*, pp. 154–158, 2016.
- [8] L. Austin and M. Thompson, "Williams [re]Mix[er]: An Interactive I Ching Composing Program," in *Proceedings of the International Computer Music Conference*, 2001.
- [9] T. Wishart, "The Composition of 'Vox-5'," *Computer Music Journal*, vol. 12, no. 4, pp. 21–27, 1988.
- [10] T. Wishart, *Red bird, A document*, Music score, U. of York Music Press, Ed., 1997.
- [11] T. Wishart, "Sonic composition in *Tongues of Fire*," *Computer Music Journal*, vol. 24, no. 2, pp. 22–30, 2020.
- [12] M. Clarke, F. Dufeu, and P. Manning, *Inside Computer Music*. Oxford University Press, 2020. Accessed: Jan. 14, 2026. [Online]. Available: <https://pure.hud.ac.uk/en/datasets/inside-computer-music-accompanying-software/>.
- [13] R. Henke, "Computermusik im Zeitalter des 'infinite undo'," in *Techno-Visionen*. Folio Verlag, 2002.
- [14] P. Théberge, *Any Sound you can imagine. Making Music / Consuming Technology*. Wesleyan University Press, 1997.
- [15] O. Baudouin, "A Reconstruction of Stria," *Computer Music Journal*, vol. 31, no. 3, 2007.
- [16] J. Rohrerhuber, A. de Campo, and R. Wieser, "Algorithms today, Notes on language design for just in time programming," in *International Computer Music Conference*, 2005, p. 291.
- [17] N. Schnell and M. Battier, "Introducing composed instruments, technical and musicological implications," in *Proceedings of the 2002 conference on New Interfaces for Musical Expression*, 2002.
- [18] P. Plessas and G. Boutard, "Transmission et interprétation de l'instrument électronique composé," *Actes des Journées Informatiques Musicale 2015, Montréal*, 2015.
- [19] Celemony Software GmbH. "10 years of ARA Audio Random Access," Accessed: Mar. 27, 2026. [Online]. Available: <https://www.celemony.com/en/10-years-of-ara>.
- [20] J.-C. Risset, "Composing in real-time?" *Contemporary Music Review*, vol. 18, no. 3, 1999.

³ For example. an undo history might record three subsequent gain changes of +5, -3 and +2dB, but will not indicate that the overall level has been raised by 4dB.

- [21] Composers Desktop Project. “CDP History: An Evolving Project,” Accessed: Jan. 21, 2026. [Online]. Available: <https://www.composersdesktop.com/history.html>.
- [22] The SoX community. “SoX - Sound eXchange,” Accessed: Jan. 22, 2026. [Online]. Available: <https://sourceforge.net/projects/sox/>.
- [23] W. Murch, *In the blink of an eye*. Silman-James Press, 2001.
- [24] G. A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for processing information.,” *Psychological review*, vol. 63, no. 2, p. 81, 1956.
- [25] T. Wishart, *Audible Design*. Orpheus the Pantomime Ltd., 1994.
- [26] G. S. Adams, B. A. Converse, A. H. Hales, and L. E. Klotz, “People systematically overlook subtractive changes,” *Nature*, vol. 592, no. 7853, pp. 258–261, 2021.

THE “RAGION TROPICA” TACTICS: INHABITING THE VIBRATION, AUDIO CARTOGRAPHY, SOUND ROUTES, AND OTHER USELESS HACKING TOOLS

Giuseppe Silvi

Conservatoire “A.Casella” of L’Aquila

grammaton@me.com

ABSTRACT

Feedback-augmented instruments unveil unmapped sonic territory. Nevertheless, cartographic tactics can trace sound routes that notation renders repeatable. This paper documents practices that emerged from augmenting an orchestral timpani with electromagnetic feedback: geometric spatial discretisation, exhaustive cataloguing of tridimensional samples, numerical analysis, and graph plots yield maps and performative notation. Each phase arose through sustained engagement with the instrument’s particular constraints—cataloguing revealed analytical needs, analysis suggested notational possibilities, notation exposed cataloguing gaps—a circular process termed *ragion tropica* (tropic reason), which operates through tactical circulation, guided by the material itself. The graphs that emerge from this circular process—where nodes represent catalogued sounds and edges encode timbral proximity in descriptor space—function as sound routes: performative trajectories through organised territory that the composer traces, producing frameworks and scores which preserve navigational flexibility, sustaining the performer’s agency within the very constraints that make compositional discourse possible.

1. INTRODUCTION

Self-designing musical instruments reshape practice by blending invention and performance, changing maker-tool-concept links. In the act of making, an ethical dimension [1] of epistemic manipulation persists: the hand that shapes its material produces knowledge exceeding the object it yields. Within a historically informed *practice*,¹ the continuous redefinition of instruments forms a permanent spine of thought. The reshaping of tools and the knowledge born from their making converge, each becoming the condition for the other’s renewal.

In 2022, I augmented an orchestral timpani through electromagnetic actuation [2, 3], allowing it to vibrate through gestural induction without physical contact and producing

¹ The term echoes but deliberately extends the notion of Historically Informed Performance (HIP): here, historical, scientific, and philosophical awareness informs the interpretation of repertoire and the practice of instrument-making and electroacoustic experimentation itself.

timbral variations unattainable with traditional techniques. However, this expanded capability revealed uncharted territory: the absence of a controllable system, the lack of methods for prescribing gestures, and the absence of an established basis for compositional organisation.

Between 2022 and 2025, I developed systematic methods: discretising the instrumental domain into spatial categories, cataloguing three-dimensional audio samples, extracting numerical data using descriptors, constructing navigable maps, and designing notation systems. These methods, which preceded and motivated the doctoral project *canto alla durata*,² are now being formalised within that research framework. What follows documents this trajectory as it took shape: through sudden laboratory discoveries and gradual refinement over time. I refer to this practice as *ragion tropica* (tropic reason) [5, 6]: a tactical circumambulation.

How can the sonic capabilities of electroacoustically enhanced instruments be organised into clear, repeatable methods for musical composition?

These organisational questions emerged within a broader inquiry into musical experience; the focus here is on the notational and cartographic dimensions of that research, resonating with Giorgio Netti’s [7] persistent question: *what is an instrument?*

My attempt to answer that question passes through practice: the laboratory becomes the site where the instrument is explored, an epistemic thing whose behaviour is not yet fully determined, and the question expands: *what is an electroacoustic instrument? An augmented one?* The work presented here is grounded in workshop experimental practice.

2. AN INSTRUMENT’S PLACES

The laboratory practice at the core of this work has been conducted since 2022 at LEAP (Laboratorio ElettroAcustico Permanente),³ a permanent electroacoustic workshop based in Rome. The *Tempo* (Timpani Electro-Magnetic Pulse Oscillation) is the case study through which these practices are documented.

² Doctoral programme in Artistic Research in Music, Conservatory of L’Aquila (DOT24BX87A, DM 629 University/Afam: CUP D11I24001070006). The project title *canto alla durata* (song to duration) derives from Peter Handke’s *Gedicht an die Dauer* [4].

³ www.leaphz.net



Figure 1. The *Tempo* instrument at LEAP laboratory: Adams Revolution 29” hand-hammered copper timpani with tetrahedral microphone array, neodymium magnet pairs on the membrane, and geometric tracings of the *Due Lune* map. On the left, the SSL BiG SiX mixer and the power amplifier driving the induction system; on the right, a S.T.ONE tetrahedral loudspeaker.

2.1 The *Tempo* Instrument

The organisational challenge arose directly from the instrument’s technical capabilities: once electromagnetic fields enable contactless vibration at any membrane point and support infinite gestural combinations, how does one move from infinite technical possibility to organised musical territory? The threshold of a musical instrument (understood as a device yielding predictable results for given actions) requires a reference system. A tactical pattern gradually revealed itself: spatial discretisation (transforming the continuous membrane surface into discrete positions), exhaustive cataloguing (systematic documentation of sonic possibilities), numerical analysis (reducing acoustic complexity to navigable parameters), and graph-based notation (constructing performative frameworks). This cartographic approach operates through controlled reduction, making the instrument’s expressive potential accessible, repeatable, and communicable: essential prerequisites for compositional discourse.

The *Tempo* instrument achieves contactless excitation through two electromagnetic inductors, mounted on wearable gloves, operating on the same principle as conventional loudspeaker drivers: audio signals drive coils, generating magnetic fields that, when brought near ferromagnetic targets, induce vibration. Rather than activating a loudspeaker cone, these fields act upon neodymium magnets anchored to the timpani membrane in pairs: each magnet on the upper surface is coupled to a twin below, sandwiching the membrane between them and fixing the assembly at a precise position.

The magnets occupy four discrete positions on the membrane, determined through geometric construction, establishing the spatial reference system that grounds the discretisation phase generating six distinct instrumental con-

figurations, each characterised by unique acoustic coupling between membrane zones and electromagnetic activation patterns.

Each configuration binds together a gestural repertoire, an acoustic space, and its electroacoustic extension—three interdependent dimensions that the following sections examine in turn.

2.2 Gesture

Sound production occurs through controlled proximity between wearable inductors and membrane-mounted magnets. The gesture of bringing the hands together—moving the inductors towards the magnets—initiates or modulates vibration. Distance governs amplitude: closer proximity increases induced vibration intensity, while greater distance attenuates or extinguishes feedback.

Contactless excitation differs fundamentally from traditional beater techniques: the electromagnetic field activates vibration from silence without percussive attack, sustains indefinitely through gestural maintenance, and develops dynamic contours impossible with struck techniques—opening access to the membrane’s sustained vibrational modes. Pedal control introduces a second gestural dimension, allowing continuous pitch variation during sustained sounds.

The combination of inductor positioning (selecting membrane activation zones), distance modulation (controlling amplitude and feedback intensity), and pedal manipulation (varying membrane tension and thus pitch) generates a three-dimensional gestural space. Navigating this space requires systematic organisation—hence the necessity of discretisation, cataloguing, and notational frameworks that gradually emerged as the *region tropica* approach.

2.3 Acoustic Space

The *Tempo* is built upon an Adams Revolution 29” hand-hammered copper timpani, with fine-tuning system, as its acoustic foundation. I selected this configuration after a three-month evaluation examining diameter, materials, and mechanical design: the 29” hand-hammered copper with a low-pedal removable mechanism offered the optimal balance of dynamic response, timbral richness, and gestural control, while the fine-tuning mechanism enables micro-adjustments for precise intonation—crucial when working with feedback-generated pitches.

I arrived at this choice through a process of elimination: tests yielded encouraging results across different bowl materials, including fiberglass models; however, high-pedal mechanisms proved impractical for the slow glissandi that electromagnetic feedback demands. The low-pedal design allows continuous, controlled pitch variation without the abrupt mechanical resistance that high-pedal systems impose at the extremes of their range. Economic constraints further narrowed the choice: professional-grade instruments were excluded in favour of a configuration whose cost remains compatible with project replication.

The current project employs a single timpani. The electroacoustic chain, described in the following section, is

entirely analogue and modular: each channel strip operates independently, making the system scalable to a full orchestral set without architectural redesign.

2.4 Electroacoustic Space

Fig. 2 shows the complete signal path. Four microphones, positioned to form a regular tetrahedron around the membrane's centre, capture vibrations and provide a three-dimensional acoustic representation of the membrane's vibrational modes. The tetrahedral arrangement supports spatial analysis of sound distribution across the surface, revealing how electromagnetic excitation couples differently to different membrane regions. The four analogue-processed channels drive the electromagnetic inductors, closing the loop: membrane vibration → microphone capture → analogue signal processing → electromagnetic excitation → membrane vibration. The recursive nature of this configuration is the instrument's defining characteristic, producing self-sustaining oscillation and gestural control of feedback behaviour [8].

Each inductor consists of a parallel network of two symmetric branches—each placing a halogen lamp in series with two magnetic coils—driven by a power amplifier's speaker output (Fig. 3). The coils generate the electromagnetic field that activates the membrane's neodymium magnets. The circuit behaves as a dynamic low-pass filter: the inductors' impedance rises with frequency, while the halogen filament's resistance increases with temperature, simultaneously compressing amplitude and shifting the spectral balance toward higher frequencies. This temperature dependence makes the inductor a non-linear, signal-dependent system whose timbral contribution changes with the signal passing through it—an active participant in the feedback loop. The parallel combination yields a total DC impedance of approximately $5.5\ \Omega$, compatible with standard power amplifier speaker outputs.

The signal chain diagram (Fig. 2) and the inductor schematic (Fig. 3) constitute two levels of electroacoustic writing: the first notates the flow of audio signals through the system; the second notates the physical mechanism by which electrical current becomes mechanical vibration. Both are forms of prescriptive notation—they specify behaviour, enable replication, and remain open to interpretation through component selection and tuning. The electroacoustic chain writes vibration onto the membrane; the circuit diagram writes the design that makes this writing possible.

Once the electroacoustic system is established and its behaviour documented, the question shifts from how vibration is produced to how its sonic results can be organised—the cartographic work that the following section details.

3. THE CARTOGRAPHIC METHOD

The cartographic method traces a directed chain—catalogue, analysis, graph, annotated graph, framework, agenda—where each phase crystallises the output of the preceding one: the catalogue's combinatorial richness is focused by analysis, structured by graph computation, annotated with

compositional signs, temporally ordered by the framework, and fixed in the score.

3.1 Spatial Discretisation

The reference system that the instrument required emerged through geometric necessity. Using four magnetic positions derived from Hippocrates' lunes (Fig. 4), I generated six distinct instrumental configurations that provide the spatial framework for subsequent cataloguing work.

The geometric discretisation of the membrane surface derives from the classical problem of Hippocrates' lunes, two intersecting circular arcs that define calculable curvilinear regions. Applied to the timpani membrane, this construction defines four magnetic constraints, denoted α , β , γ , δ , relative to the centre point N (North), as shown in Fig. 4.

Each neodymium sandwich pair grips the membrane through mutual magnetic attraction (approximately $6Kg$ pull force per pair), holding its position without mechanical fastening. Sliding the pair along the surface with sufficient force allows repositioning at any of the four nodes. The six possible pairings define six distinct instrumental configurations:

- *Position I:* $\alpha\beta$ (tight spacing between magnets)
- *Position II:* $\beta\gamma$ (tight spacing)
- *Position III:* $\gamma\delta$ (tight spacing)
- *Position IV:* $\alpha\gamma$ (intermediate spacing)
- *Position V:* $\beta\delta$ (intermediate spacing)
- *Position VI:* $\alpha\delta$ (wide spacing)

The nomenclature follows ordinal distance logic—positions with closer magnetic spacing (I–III) precede those with intermediate (IV–V) and maximum (VI) spacing—transforming the continuous membrane surface into a navigable, discrete territory. Each position produces distinct acoustic coupling between membrane zones and electromagnetic activation patterns (Fig. 7), yielding characteristic timbral behaviours (Fig. 10).

Two intersecting arcs traced on the membrane surface establish the spatial reference system; magnetic constraints materialise as positioned inductors at calculated coordinates. The physical inscription embodies the transition from geometric abstraction to operational territory—from theory to practice (Figs. 5, 6).

3.2 Catalogue

With six spatial positions defined, the cataloguing process methodically addressed the combinatorial space. For each position, I documented 72 quadraphonic samples along two control dimensions.

Firstly, the pedal tension (12 positions, 0–11): from minimal to maximum membrane tension, producing continuous pitch variation across approximately one octave per position. Pedal positions were indexed numerically for precise replication across sessions.

Secondly, the inductor configuration (6 possibilities): single-inductor activations ($\alpha(A)$, $\gamma(A)$, $\alpha(C)$, $\gamma(C)$) plus two-inductor combinations ($\alpha(A)\gamma(C)$, $\gamma(A)\alpha(C)$) (here exemplified for Position IV). The two-inductor configurations produce “electromagnetic chords”—simultaneous

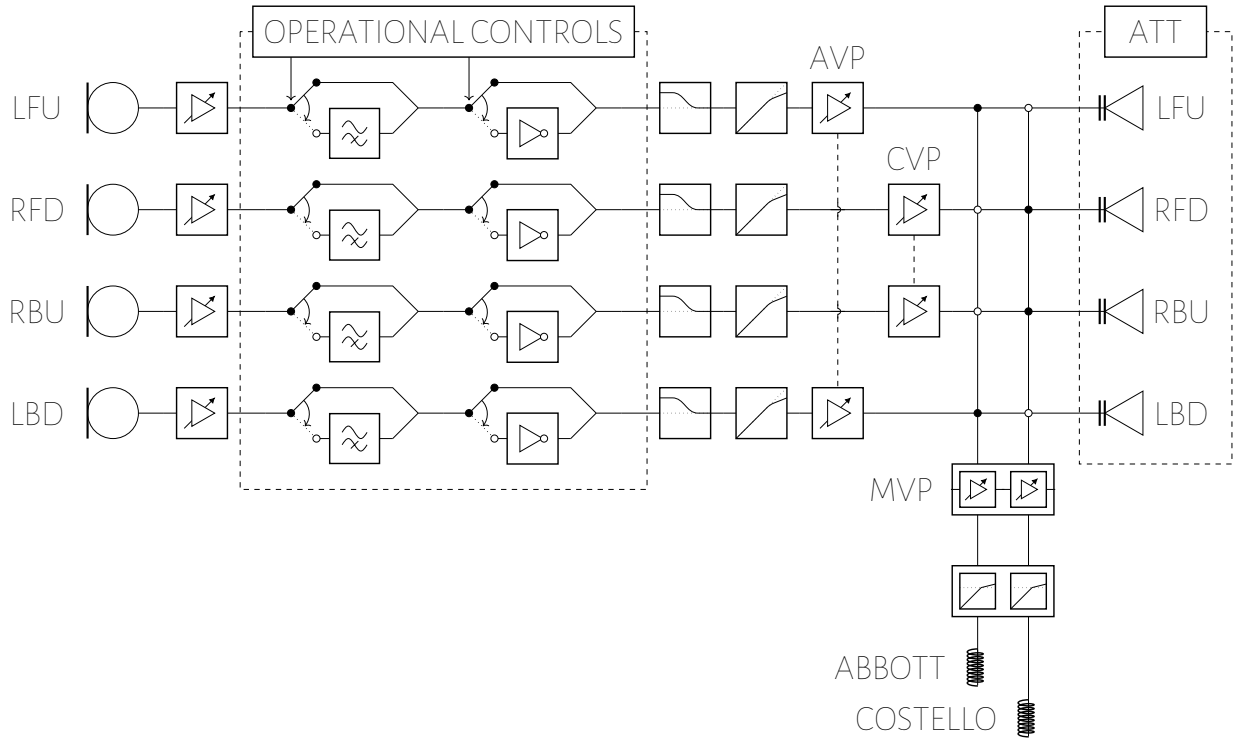


Figure 2. Electroacoustic analogue signal chain of the *Tempo* instrument. Four omnidirectional microphones, in a tetrahedral spaced configuration, capture membrane vibrations. An SSL BiG SiX analogue mixer processes each channel through three-band equalisation (LF shelf at 60 Hz boosted to compensate inductor roll-off) and channel compression (limiting feedback peaks); the G-Series bus compressor provides a second limiting stage on the summed signal before it drives electromagnetic inductors (*Abbott* and *Costello*) positioned on the membrane surface, creating recursive feedback. AVP (*Abbott* Volume Pedal) and CVP (*Costello* Volume Pedal) control individual inductor intensity via expression pedals; MVP (Master Volume Pedal) governs the overall feedback level. ATT (*Amplificazione Tetraedrica Trasparente*) provides tetrahedral sound reinforcement through spherical loudspeakers (S.T.ONE)—a predisposition, not a requirement: the instrument is entirely analogue and self-contained. The BiG SiX USB interface enables 96 kHz digital capture for cataloguing purposes.

activation of two membrane zones generating interference patterns and coupled resonances.

The resulting 72 samples per position ($12 \times 6 = 72$), multiplied across six positions, yield 432 documented sounds—the complete catalogue of the *Tempo* instrument’s systematically organised sonic territory (Fig. 9). The waveform displays (Fig. 8) reveal spatial signatures across the four channels: sample $7\gamma(C)$ shows balanced energy distribution, while $11\alpha(C)$ concentrates energy in channels 1 and 3, distinguishing distributed from membrane-localised responses.

The catalogue functions as constitutive documentation [9]: the act of cataloguing shapes understanding of the material, revealing patterns and necessitating analytical frameworks.

My cataloguing was non-linear and reiterative. I re-catalogued each position as tuning procedures evolved from exploratory tunings towards the systematic *D3*-referenced protocol documented in Sec. 4.1. I catalogued the first three positions (Studies I-III) with less refined tuning; Position IV was the first generated with a consolidated tuning method. The temporal dimension—the catalogue as a document of evolving technical mastery—remains legible in the materials: early catalogues exhibit greater pitch variability, while later catalogues show tighter intonation control.

The catalogue numbers (Fig. 9) encode this systematic-

sation: each sample identifier specifies position, inductor configuration, and pedal state, for precise recall and replication.

3.3 Analysis

The 72 quadraphonic samples per position required reduction to a navigable form. Raw audio provides too much information: waveforms, spectral content, spatial distribution, temporal evolution—all simultaneously present, all relevant, all overwhelming when multiplied across hundreds of samples.

I implemented numerical analysis in Wolfram Mathematica to extract quantitative descriptors from each sample.

```
SetOptions[SelectedNotebook[],
  PrintPrecision -> 5];
files = FileNames["*.wav",
  NotebookDirectory[]];
collection = Import[#] & /@ files;
qualities = AudioMeasurements[
  collection,
  {"SpectralSpread",
    "LoudnessEBU",
    "SpectralRollOff"},
  "List"];
```

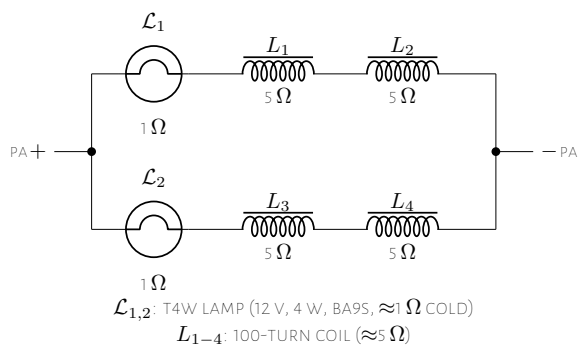


Figure 3. Circuit diagram of the electromagnetic inductor network (mk01) for *Abbott and Costello*—named after the heptapod linguists in Denis Villeneuve’s *Arrival* (2016), whose cavernous vocalisations the first sounds produced by the instrument recalled. Two symmetric branches share the power amplifier output: each branch places a halogen lamp ($\mathcal{L}_{1,2}$) in series with two coils (L_{1-4}). The schematic operates as a form of circuit transcription, prescribing electron flow as a score prescribes sonic behaviour.

Three descriptors turned out to be particularly useful for distinguishing timbral territories:

Spectral Spread measures the bandwidth of spectral content—how widely energy is distributed across the frequency spectrum. A narrow spread indicates focused spectral energy (near-sinusoidal behaviour); a wide spread suggests rich harmonic or inharmonic content. Spectral Spread distinguished membrane-localised sounds (narrower spread) from bowl-resonant behaviours (wider spread due to coupled resonances).

Loudness EBU quantifies intensity using the European Broadcasting Union standard (measured in LUFS—Loudness Units relative to Full Scale). Standardised measurement enables comparison across samples recorded in different sessions, accounting for variations in electromagnetic coupling efficiency and membrane response. Loudness values revealed how pedal tension affects vibration amplitude.

Spectral Roll-Off identifies the frequency below which a specified percentage (typically 85%) of spectral energy concentrates. Roll-Off captures the distinction between bass-heavy membrane vibrations and brighter bowl resonances. Roll-off values also showed that two-inductor configurations produce different spectral balances from single-inductor configurations.

The analysis produced a dataset organising the 72 samples for each position in three-dimensional descriptor space (Spectral Spread, Loudness EBU, Spectral Roll-Off). Numerical reduction transformed acoustic complexity into navigable coordinates—each sample now occupies a specific location in this abstract space, with proximity indicating timbral similarity.

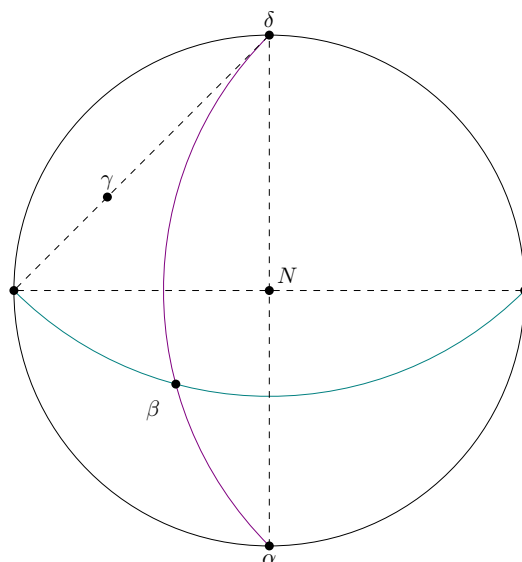


Figure 4. Theoretical map *Due Lune, Ippocrate su pelle* (Two Moons, Hippocrates on skin). Spatial discretisation of the membrane surface determines four magnetic constraints ($\alpha, \beta, \gamma, \delta$) in relation to the centre point N (North). The map generates six positions, each with its own tuning.

These three descriptors emerged through pragmatic experimentation: other candidates (spectral centroid, harmonic-to-noise ratio, spectral flux) provided less effective differentiation for this instrument’s behaviours.

3.4 Graph

```
FindSpanningTree[
  qualities,
  PlotTheme -> "Monochrome",
  VertexLabels -> Placed[
    {"Index"}, {Center}],
  VertexSize -> .5]
```

The three-dimensional descriptor space (Spectral Spread, Loudness EBU, Spectral Roll-Off) provided the basis for graph construction. Using Mathematica’s spring-embedded layout algorithms, I generated graphs where:

- *Nodes* represent individual catalogued samples
- *Edges* connect samples that are proximate in descriptor space
- *Spatial layout* reflects timbral relationships: proximate nodes sound similar, distant nodes sound distinct

The spring-embedded algorithm treats nodes as physical objects connected by springs—strong connections (similar descriptors) pull nodes together, while dissimilar nodes repel each other. The algorithm iterates until a stable configuration is reached, producing two-dimensional visualisations of three-dimensional proximity relationships.

Six graphs emerged, one per position (see example for Position IV in Fig. 11). Each graph reveals a distinct topological structure reflecting that position’s timbral territory. Some graphs exhibit apparent clustering—groups of tightly

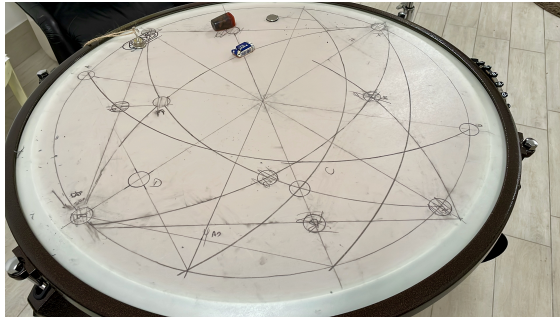


Figure 5. Exploratory geometric constructions pencilled on the timpani membrane during the search for the definitive *Due Lune* configuration. Successive attempts to project the theoretical cartography onto the physical surface accumulate as overlapping arcs, intersections, and candidate constraint points—a palimpsest of the discretisation process.



Figure 6. Top view of the freshly traced *Due Lune* (Two Moons) cartography on the membrane, with neodymium magnets still off-position. The four constraint points arise from distinct geometric intersections: α from lune and circumference, β between the two lunes, γ on an external triangular intersection axis, δ from circumference and lune—replicating the theoretical map of Fig. 4.

connected nodes representing timbral families—while others show more distributed structures suggesting continuous gradation.

These topological differences reflect fundamental acoustic distinctions between positions. Position I, with tight magnetic spacing $\alpha\beta$, generates relatively homogeneous responses—the graph shows few distinct clusters. Position VI, with maximum spacing $\alpha\delta$, activates a more diverse set of membrane zones—the graph exhibits pronounced clustering, with each cluster corresponding to a distinct coupling mode between the membrane and bowl.

The graphs function as navigational tools: pathways through the graph suggest performative trajectories—sequences of sounds connected by timbral proximity. Edges between nodes generate smooth timbral evolution; jumping between disconnected clusters produces abrupt timbral shifts. The graph structure thus encodes both similarity (connected nodes) and contrast (disconnected regions), pro-

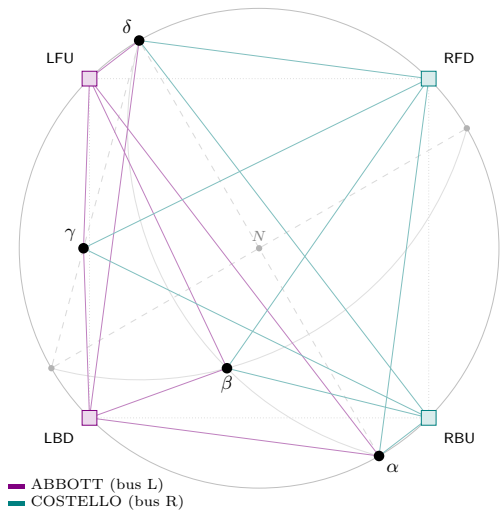


Figure 7. Microphone placement on the *Due Lune* map. The lune pattern is rotated to match the actual membrane orientation—a rotation found empirically through successive attempts to optimise the microphone-to-map relationship. Rays connecting each microphone to the induction nodes reveal the varying distances that shape distinct feedback modes across the six instrumental positions.

viding a compositional vocabulary that preserves performer agency.

Within the directed chain, the graph marks a pivot: the open analytical potential accumulated through cataloguing and descriptor extraction begins here to solidify into compositional structure.

3.5 Annotated Graph

The graph is an analytical object: its topology is fixed by computation and descriptor proximity. Annotation transforms it into a compositional tool—a spatial instrument for studying trajectories, exploring deviations from ordinary paths, and composing itineraries through timbral territory. The temporal dimension remains absent: it is the framework that converts spatial routes into scored time.

Fig. 11 shows the annotated graph for Position IV. Each node receives multiple annotations:

Sample index The catalogue number for precise recall (e.g., sample 23 = pedal position 3, inductor configuration $\alpha(A)\gamma(C)$);

Pitch notation The fundamental frequency in cent-precise notation at an A=442Hz reference;

Timbral character Brief descriptive labels emerging from listening—“membrane”, “bowl”, “coupled”, “unstable”—functioning as performative reminders.

The annotation process revealed that graphs function as maps requiring keys—the topological structure provides spatial relationships, but performers need concrete referents

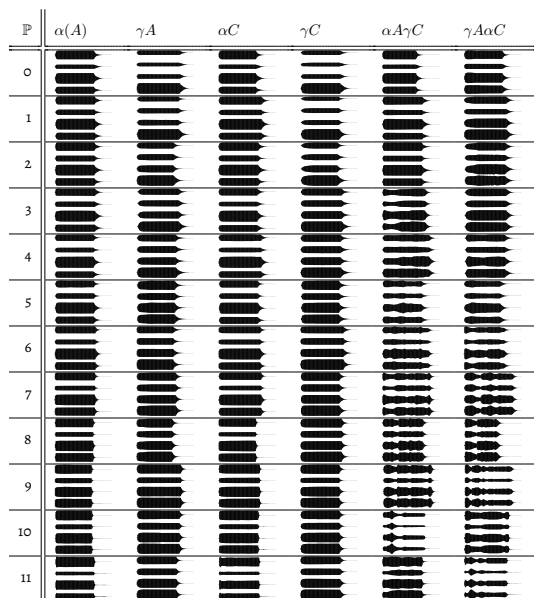


Figure 8. Catalogue of waveforms for the 72 quadraphonic samples. Each cell displays four waveforms (tetrahedral microphone array). Rows index pedal tension positions (P : 0 – 11); columns indicate inductor configurations: $\alpha(A)$, $\gamma(A)$, $\alpha(C)$, $\gamma(C)$, and two-inductor combinations $\alpha(A)\gamma(C)$, $\gamma(A)\alpha(C)$.

to navigate the territory. The sample indices connect abstract nodes to specific sounds in the catalogue; the pitch notations provide melodic reference; the character labels offer timbral orientation.

Pathways through the annotated graph suggest compositional strategies. Following connected nodes generates smooth timbral progressions—each transition moves to a neighbouring sound in descriptor space, producing gradual evolution. I trace routes through the graph: node 12 → 15 → 23 → 31 defines a specific timbral trajectory, a composed pathway through the instrument’s organised sonic territory.

The graphs preserve agency: they organise possibilities while leaving outcomes open. Multiple pathways connect any two nodes; disconnected regions invite sudden jumps; cyclic paths enable returns and variations. The notation captures relationships—this sounds like its neighbours. The performer navigates the map; the map does not choreograph the performance. The structure resonates with Borges’ labyrinths [10]: maps that preserve the possibility of getting lost while providing orientation.

3.6 Framework

The annotated graphs organise sonic territory spatially but require temporal instantiation for performance. The need gave rise to the *telaio* (framework)—a notational structure that transcribes spatial graphs into temporal sequences while preserving navigational flexibility.

Fig. 12 shows the framework for Position IV’s Study. The notation operates through multiple simultaneous layers:

Figure 9. Systematic catalogue of the 72 samples for position IV. Columns: INDEX (progressive numbering), CODICE (audio file identifier), ABBOTT/COSTELLO (inductor positions: α or γ in configurations A or C), PEDALE (tension pedal positions 000–110), R_{442} (pitch in cent-precise notation at A=442Hz), Hz (fundamental frequency). The catalogue organises the combinatorial space: 12 pedal positions $\times$ 6 inductor configurations = 72 samples per position.

Upper metadata strip Records the sample name (linking the framework to the catalogue), the node identifier (linking it to the graph), and a section label (intro, A, B, etc.) for structural organisation. Each vertical line represents a node (catalogued sound); horizontal reading generates a timbral contour.

Central strip Traces the pedal tension profile across the sequence, notating membrane tension contour as a continuous line.

Lower staves Indicate inductor positions and activation states. Four lines represent the four magnetic positions (α , β , γ , δ); filled circles show active inductors, empty circles inactive.

The framework is not a performance score but a compositional tool: I read it at the instrument, testing the oriented sequence of actions against the instrument’s actual behaviour. Choices made at the annotated graph stage are transcribed into the framework and explored in sequence, allowing me to understand how the material unfolds in time—how gesture relates to duration through the instrument’s feedback response. The framework is thus a scaffold for writing the

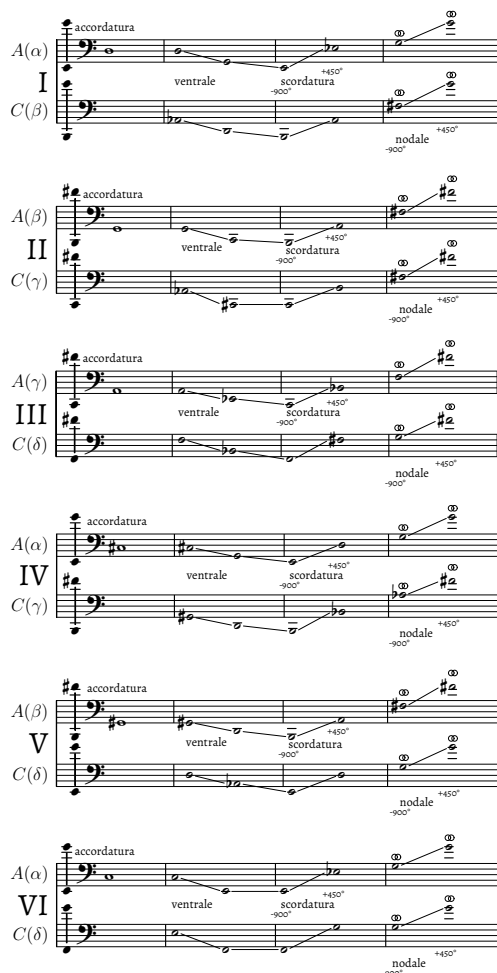


Figure 10. Timbral ranges (*ambiti*) of the six positions. Each position displays: *accordatura* (fundamental reference pitch), *ventrale* trajectory toward membrane centre (belly region, $-900^\circ/+450^\circ$ cent deviations), *nodale* trajectory toward nodal regions, for both inductors *Abbott* and *Costello*. The systematic documentation reveals how membrane tension couples with electromagnetic positioning to generate distinct sonic territories.

agenda: it is where I learn from the instrument before the final score crystallises.

3.7 Agenda

The frameworks (*telai*) for each position provide navigational structures yet require integration into coherent performative works. I composed the *Sei Studi di Agamotto sul Tempo* (Six Studies of Agamotto on Tempo) as the final stage (Fig. 13)—six studies that function simultaneously as a pedagogical progression, a technical exploration, and musical compositions.

The term *agenda* (from Latin *agere*, to do/act) designates these performative scores as action-oriented. This orientation resonates with recent calls to develop notation conventions that acknowledge the role of technology as a constitutive element of augmented instrument works [11].

Each Study organises its framework materials according

to distinct compositional strategies:

Studies I–III (*Studi Tecnici*, Technical Studies) systematically explore the progressive technical complexity: from fixed electromagnetic output (Study I) through unified dynamic control (Study II) to independent inductor control (Study III).

Studies IV–VI (*Studi Espressivi*, Expressive Studies) transition towards musical expression, applying technical mastery to compositional invention. These studies employ graph-based frameworks with greater freedom: longer pathways through descriptor space, more complex timbral trajectories, and attention to large-scale formal arcs.

The Studies also reveal temporal layering: Studies I–III were composed in 2022, Study IV in spring 2024; Studies V–VI are currently in progress. The chronological sequence documents a learning trajectory—the Studies preserve the progressive mastery of the instrument’s capabilities through their increasing compositional ambition. Alongside Studies V–VI, I am developing a solo work that moves beyond the position-by-position structure of the Studies: a piece that draws on the full range of possibilities revealed across all six positions, integrating the instrument’s capabilities into a unified compositional argument.

4. TRANSFORMATIVE PRACTICES

The cartographic chain presupposes practices of a different order: the operational work of tuning and stabilising *Tempo* for systematic exploration, and the epistemic positioning of augmentation within a broader tradition of transformative practices.

4.1 Operational Practices

I developed the tuning practice iteratively, moving from exploratory “travelling” tunings towards systematic, replicable procedures.⁴ The current standardised tuning protocol establishes Position I as the reference configuration: inductor *Abbott* activating magnet α , pedal tension at maximum (position 11), fine-tuning set to zero degrees, yielding $D3$ at $442Hz$ concert pitch. Investigating membrane-bowl coupling revealed explosive resonance around $D3$ after acoustic annihilation around $C\sharp3$: this discovered phenomenon provides the reference pitch for the entire tuning system. From this $D3$ foundation in Position I, the fine-tuning mechanism extends the tuning across the six positions, covering a comprehensive pitch range. (Fig. 10)

Tuning stability requires careful attention to mechanical detail. The membrane must undergo repeated cycles of tension and release during setup to stabilise its molecular structure. Individual tension rods require individual verification: counting exposed thread turns for each bolt ensures homogeneous tension distribution around the membrane’s circumference. Any compensation adjustments require subsequent pitch verification.

Signal robustness is verified through polarity testing: rapid polarity inversions should produce an immediate feedback

⁴ Video documentation of the tuning procedures and other video about *Tempo* are available at: <https://www.youtube.com/playlist?list=PLLfETWgVXqYaoOSBXed7lGm2BG77-QHhL>

CAD IV dr

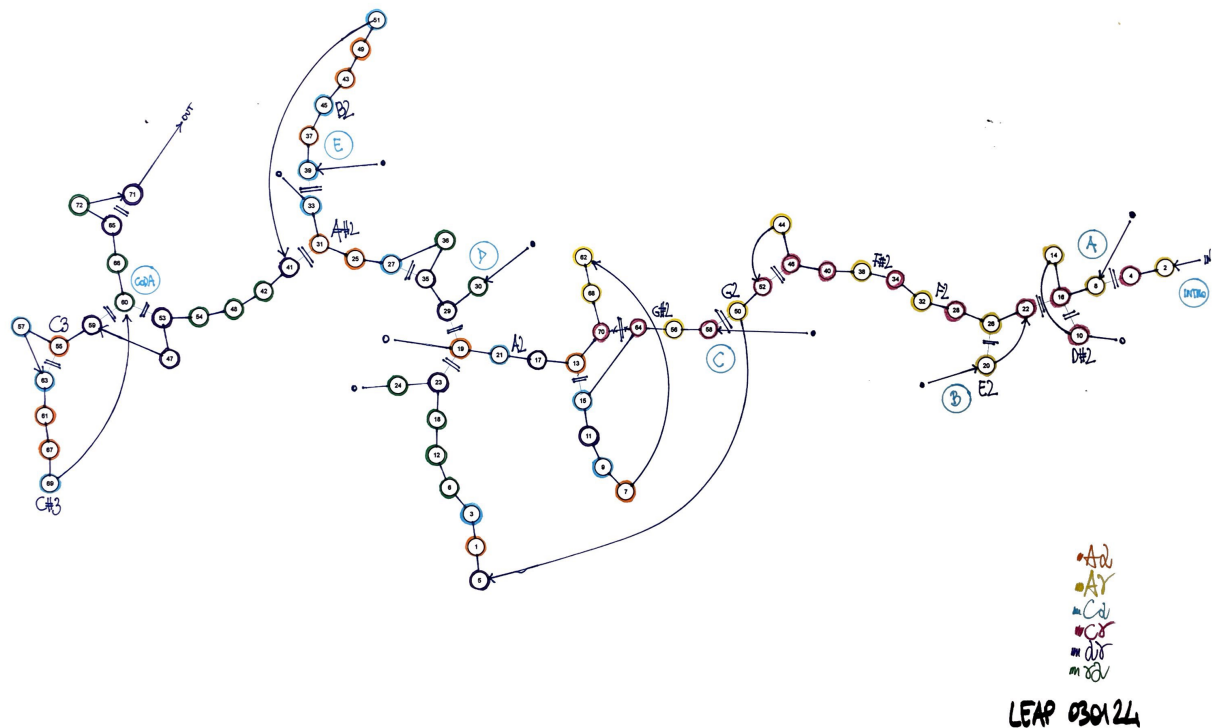


Figure 11. Annotated graph of position IV. Each node represents a catalogued sample; connections indicate three-dimensional proximity in descriptor space (Spectral Spread, Loudness EBU, Spectral Roll-Off). The marked pathway traces a composed route through the timbral territory: the entry point is arbitrary—a compositional choice that determines where the study begins. Other graphs are read with different logics (e.g. left-to-right traversal, alternative deviation strategies). The framework (Fig. 12) transcribes the chosen route into temporal notation.

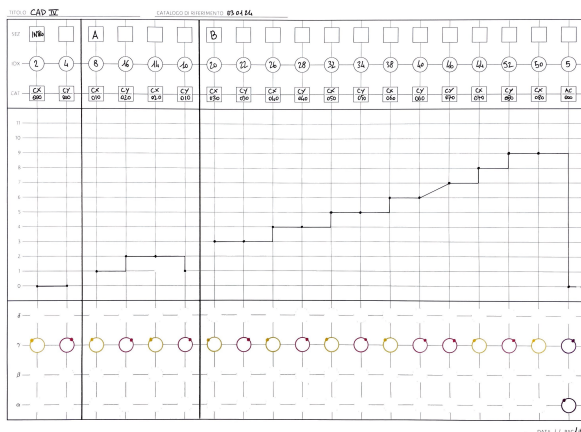


Figure 12. Gestural framework (*telaio*) for Position IV, transcribing the annotated graph (Fig. 11) into temporal notation. Upper metadata strip: sample name, graph node identifier, and section label (intro, A, B...). Central strip: pedal tension profile as a continuous contour. Lower staves: four lines for magnetic positions (α , β , γ , δ); filled circles indicate active inductors, empty circles inactive. Each vertical alignment represents one catalogued sound; horizontal reading generates the timbral trajectory of the study.

response at the lowest available pitch for the current configuration. Engaging the high-pass filter eliminates out-of-band signals that could destabilise the feedback system. These verification procedures, developed through concert experience (particularly the LAZZARO project concert series 2023-2024),⁵ evolved from hour-long pre-concert preparations into rapid, reliable setup protocols.

4.2 Epistemic Practices

The practice of *preparation*, temporary physical modification of acoustic instruments, constitutes an established tradition from Cage to Netti, evolving from idiomatic exploration to ontological inquiry: each preparation defines a sonic world; the score navigates this world through controlled variations.

The electroacoustic *amplification* operates as ontological renewal, illuminating thresholds once unimaginable or private.

The *hacking* practice denotes creative tampering through unorthodox access to instrumental mechanisms, privileging intuition-driven exploration over systematic methodology. In this progression, I define *augmentation* as the

⁵ www.leaphz.net/lazzaro

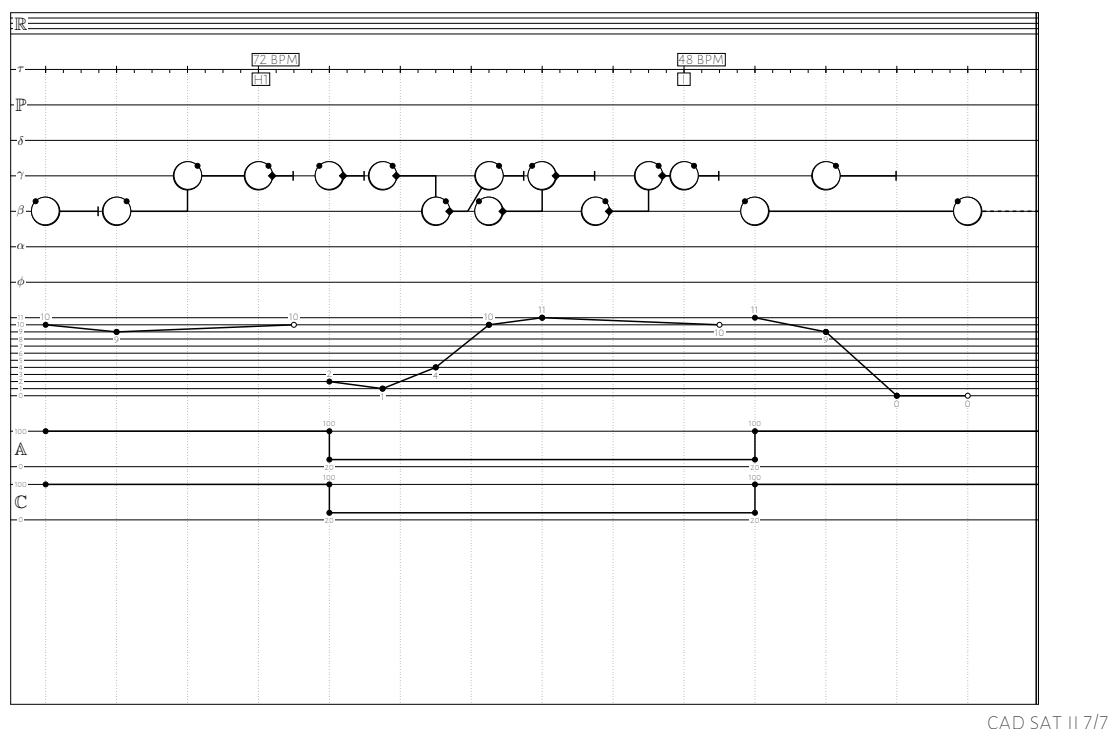


Figure 13. Final page of the “score-agenda” (*partitura-agenda*), *Secondo Studio di Agamotto sul Tempo* (Second Study of Agamotto on Tempo). The upper staff, marked $\mathbb{R}$, is reserved for pitch transcription in traditional notation—a predisposition for communicating in conventional musical language, unused in this study but available for future works requiring it. Below, the framework (*telaio*) layers encode the gestural and electromagnetic structure of the performance.

extension of a traditional instrument’s expressive possibilities through electroacoustic complement, while preserving its physical principles and historical memory [12, 13]. The acoustic instrument retains its primary function and technical-expressive heritage; the electroacoustic system adds modes of excitation and articulation that are impossible with traditional techniques alone.

Augmentation synthesises aspects of previous practices—the systematic approach of preparation, the sonic revelation of amplification, the exploratory spirit of hacking—while introducing a crucial element: *recursive feedback* between the electroacoustic system and the acoustic resonator [3]. In my practice, recursive feedback is the operative principle that runs through hacking, augmentation, composition, and performance: it is at once the technical lever that enables contactless excitation, the force that demands new notational strategies, and the medium through which instruments and environment interact in concert [14, 15].

5. CONCLUSIONS

Inhabiting vibration demands cartography; cartography yields routes; routes require tools whose value remains bound to the territory that generated them.

The *racion tropica* tactics documented here trace a di-

rected chain through which the *Tempo*’s sonic territory became organised. Geometric discretisation of the membrane surface generated six instrumental positions; exhaustive cataloguing documented 432 quadrasonic samples across pedal and inductor configurations; numerical analysis reduced acoustic complexity to three navigable descriptors; graph computation revealed timbral proximity structures; notation transcribed spatial routes into temporal scores. The method did not precede practice: it revealed itself through sustained laboratory engagement with the instrument’s particular constraints, each phase’s output becoming the next phase’s material—cataloguing revealed analytical needs, analysis suggested notational possibilities, notation exposed cataloguing gaps. This circular movement confirmed that tactical circulation [16] through material generates organisational methods.

The tools that emerged—Hippocrates’ lunes traced on skin, tetrahedral capture, spring-embedded graphs, framework notation—are deliberately termed *useless* in Mario Bertoncini’s [17] sense of *bottega d’arte*: instruments of craft-based practice that operate through circulation and borrowing. This rationality, which María Zambrano’s *razón poética* [5, 6] names as *racion tropica* (from Greek *trópos*: turn, deviation, figure), proceeds tactically through the tension between invention and writing as instruments for un-

derstanding musical experience [18, 19, 20]. The *Sei Studi di Agamotto sul Tempo* preserve this trajectory as action-oriented scores—*agenda*—whose navigational flexibility sustains the performer’s agency within compositional constraints. The work positions itself within the Roman electroacoustic tradition, responding to Franco Evangelisti’s call for *un nuovo mondo sonoro* [21] through contemporary technical means, informed by the self-design principles of Enzo Mari’s *autoprogettazione* [1]: autonomy in judgement, integration of theory and praxis through the hand that shapes its material.

The *ragion tropica* pattern emerged through the *Tempo* instrument but exists within a broader research context. The doctoral project *canto alla durata* encompasses four hacked/augmented instruments—timpani, contrabass clarinet, cello, and voice—each representing a distinct instrument family. The organisational work documented here has been fully undertaken only for the timpani; the remaining instruments exist and function, their systematic exploration at varying stages of development. Maps do not make the territory navigable for the first time—experienced improvisers navigate uncharted ground—but they enable communicability and repeatability, prerequisites for compositional discourse beyond individual exploration. The ensemble dimension represents a further organisational challenge: coordinating multiple augmented instruments across distinct sonic worlds while sustaining tactical flexibility.

Acknowledgments

With the support of the doctoral program in Artistic Research in Music, Conservatory of L’Aquila (DOT24BX87A, DM 629 University/Afam: CUP: D11I24001070006).

With the support of LEAP (Laboratorio ElettroAcustico Permanente), Rome, and of the LEAPers whose patient participation accompanied the growth of the instrument—and, through it, my own.

6. REFERENCES

- [1] E. Mari, *Autoprogettazione*. Mantova: Edizioni Corraini, 2002, ristampa 2010.
- [2] M. M. Wanderley and P. Depalle, “Gestural control of sound synthesis,” *Proceedings of the IEEE*, vol. 92, no. 4, pp. 632–644, 2004.
- [3] D. Overholt, E. Berdahl, and R. Hamilton, “Advancements in actuated musical instruments,” *Organised Sound*, vol. 16, no. 2, pp. 154–165, 2011.
- [4] P. Handke, *Gedicht an die Dauer*. Berlin: Bibliothek Suhrkamp, 1986.
- [5] M. Zambrano, *Chiari del bosco*. Milano: SE, 1991.
- [6] —, *Filosofia e poesia*. Bologna: Pendragon, 2010.
- [7] G. Netti, “Che cos’è uno strumento,” IMPULS Academy, Graz, Jul. 2023. [Online]. Available: <https://youtu.be/CoSVrkYYo3w>
- [8] A. McPherson, “The magnetic resonator piano: Electronic augmentation of an acoustic grand piano,” *Journal of New Music Research*, vol. 39, no. 3, pp. 189–202, 2010.
- [9] M. Ferraris, *Documentalità*. Roma-Bari: Editori Laterza, 2014.
- [10] J. L. Borges, *Finzioni*. Torino: Einaudi, 1955.
- [11] C. Ramos Flores and J. R. Sigal Sefchovich, “Towards common ground in notation for augmented instruments,” *Organised Sound*, pp. 1–12, 2026.
- [12] M. Lupone, L. Bianchini, S. Lanzalone, and A. Gabriele, “Research at Rome’s Centro Ricerche Musicali on interactive and adaptive installations and on augmented instruments,” *Computer Music Journal*, vol. 44, no. 2-3, pp. 133–158, 2021.
- [13] C. A. Boracchi, *Manuale pel Timpanista*. Luigi di Giacomo Pirola, 1839.
- [14] A. Di Scipio, “‘Sound is the interface’,” *Organised Sound*, vol. 8, no. 3, pp. 269–277, 2003.
- [15] —, “Listening to Yourself through the Otherself,” *Organised Sound*, vol. 16, no. 2, pp. 97–108, 2011.
- [16] M. de Certeau, *L’invenzione del quotidiano*. Roma: Edizioni Lavoro, 2010.
- [17] M. Bertoncini, *Arpe eolie e altre cose inutili / Aeolian Harps and other useless things*. Milano: Die Schachtel, 2007.
- [18] D. Guaccero, *Un iter segnato: scritti e interviste*, ser. Le sfere, A. Mastropietro, Ed. Ricordi LIM, 2005.
- [19] F. Brook, V. Minardi, and D. Andrews, Eds., *Symposium Internazionale sulla Problematica dell’Attuale Grafia Musicale*. Roma: Tipografia Bruno Cruciani, 1974, atti del convegno tenutosi presso l’ILA, Roma, 23–26 ottobre 1972.
- [20] H. Bergson, *Sul segno. Lezioni del 1902-1903 sulla Storia dell’idea di tempo*, ser. Filosofia al presente, R. Ronchi and F. Leoni, Eds. L’Aquila: Textus, 2011.
- [21] F. Evangelisti, *Dal silenzio a un nuovo mondo sonoro*, ser. Musica, I. Heimbächer Evangelisti, Ed. Roma: SEMAR, 1991, no. 8.

MASS-SPRING LATTICE PHYSICAL MODELING SYNTHESIS USING A QUANTUM ALGORITHM

Xavier Davenport

University of Illinois at Urbana-Champaign
xavierd2@illinois.edu

ABSTRACT

In this paper I describe how a recently developed quantum algorithm for simulating the dynamics of mass-spring systems can be applied in the artistic context of physical modeling synthesis, and I propose a scalable “virtual microphone” method based on amplitude amplification for extracting perceptually relevant information from the simulated system. After outlining how to synthesize sound using a one-dimensional mass-spring system with both fixed and unfixed endpoints, I discuss the scalability and implications of this approach as quantum hardware develops. Potential applications of this theoretical framework include the simulation of imperfect vibrating systems and the large-scale modeling of systems containing up to on the order of 10^{29} coupled masses in a 100 qubit system.

1. INTRODUCTION

Waveguide and modal synthesis are two of the most historically important and widely used methods of audio physical modeling. Digital waveguide methods use delay lines to model systems with high frequency losses and dispersion, and can be used to emulate the sound of strings, brass, and woodwind instruments, among others, reasonably well [1]. Modal synthesis requires knowledge of an instrument’s resonant characteristics, including how the resonances dampen over time and respond to external excitation [2]. These synthesis techniques can be executed quickly enough to model instruments in real-time on modern hardware for a wide variety of resonating structures [3].

The computational efficiency of contemporary physical acoustic physical modeling techniques is achieved through approximations and simplifications of the underlying physics of vibrating systems. For example, the Karplus-Strong waveguide synthesis model assumes an ideal string that does not naturally decay in a realistic way [4]. An artificial envelope must be applied to the sound if it is intended to sound like a plucked string. There is also generally no simple way to add imperfections to the string in a waveguide model, such as one might find in the gut strings of a lute.

A more naïve approach to modeling a string involves representing the structure as a one-dimensional lattice of masses connected to each other with springs. If the string consists

of only a single mass fixed via spring to two endpoints, then the plucked string has one resonant mode, the fundamental frequency, which is dependent upon the size of the mass and the strength of the springs. As more and more masses are added to the system, the number of resonant modes increases, and the string exhibits a realistic harmonic resonant spectrum. The challenge of simulating this mass-spring system is the sheer number of calculations required to describe the time evolution of the system. Every mass tugs on its neighbor, and its neighbor’s neighbor, and every other mass in the system, and the same is true of every other mass! The number of computational operations required to examine the net force on each mass grows very quickly as the size of the system increases.

Recent advances in computational power and software development within the past decade have made possible the simulation of mass-spring systems containing hundreds of masses in real-time [5]. In the following, I outline a method for using quantum computers to simulate 2^{n-1} masses, where n is the number of qubits available on a quantum computer. The scaling of this simulation is exponential: with a modest number of qubits, an immense number of masses in a mass-spring system can be simulated in any number of dimensions far outpacing any classical algorithm [6]. I begin with a brief overview of the relevant concepts and mathematics from quantum computing before discussing how to map the mass-spring problem to the Schrödinger equation. I then classically simulate the quantum algorithm to find the dynamics of a plucked string with unfixed and fixed endpoints while emphasizing the simple physical intuition of this approach to modeling based on my experience as a guitarist. Next, I propose a method of using amplitude amplification to measure only the velocity of a single mass to recover an audio waveform. Finally, I discuss the long-term implications of this approach for physical modeling synthesis.

2. BASIC QUANTUM COMPUTING

A quantum computer is fundamentally different from the ordinary digital computers we interact with on a daily basis. This section provides the necessary conceptual foundation and some basic intuition for why quantum computers are a good candidate for large-scale mass-spring simulation. For a deeper understanding of quantum computation, I recommend this introduction to quantum computing for computer musicians [7], this introduction to quantum computing for non physicists [8], and the seminal text on quantum com-

puting by Nielsen and Chuang [9].

The property of quantum systems most relevant to the calculations described in this paper is superposition. There is a quantum version of the bit called a qubit that can exist in a superposition of the states $|0\rangle$ and $|1\rangle$. The binary state 0 is represented by $|0\rangle$, the binary state 1 is represented by $|1\rangle$, and they can each be represented using matrices in the following way.

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (1)$$

An arbitrary superimposed state $|\psi\rangle$ can be written as a linear combination of these two states in the form

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (2)$$

When the qubit is measured at the end of a calculation, the state collapses into either $|0\rangle$ with probability α^2 or $|1\rangle$ with probability β^2 . The normalization condition $|\alpha|^2 + |\beta|^2 = 1$ ensures these probabilities sum to one.

The computational power of quantum systems becomes apparent when qubits are combined to form multi-qubit states. For two arbitrary states $|\psi_1\rangle$ and $|\psi_2\rangle$, their combined state can be made using a tensor product in the following way.

$$|\psi_1\rangle \otimes |\psi_2\rangle = \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} \otimes \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \alpha_1\alpha_2 \\ \alpha_1\beta_2 \\ \beta_1\alpha_2 \\ \beta_1\beta_2 \end{pmatrix} \quad (3)$$

This sort of combined state is often written in shorthand to omit the tensor product symbol as $|\psi_1\rangle \otimes |\psi_2\rangle = |\psi_1\psi_2\rangle$. For every qubit added, the number of possible output states doubles. In other words, the dimension of the state vector grows exponentially with the number of qubits. For example, if I am working with three qubits, then there are eight possible outcomes when I measure the system (e.g., $|000\rangle, |001\rangle, |010\rangle, \dots$ etc.). The simplest way to reconstruct the probability distribution across these states requires many measurements of identically prepared systems. A large part of research surrounding quantum algorithms is (1) developing clever methods of creating a superimposed state then (2) efficiently extracting information from it. As I discuss in the next section, there happens to be an effective way to describe the energies of masses in a mass-spring system where each possible measurement outcome corresponds to a mass velocity. Later in this paper, I discuss how to use amplitude amplification to efficiently extract information from the system.

3. MASS-SPRING HAMILTONIAN SIMULATION

In a mass spring-system, the Newtonian force $F = ma$ is equal to the restoring force defined in Hooke's law, $F = -kx$. The motion of a single mass connected to a single spring fixed at the opposite end can be described using

$$ma = -kx, \quad (4)$$

where m is mass, a is acceleration of the mass, k is a spring constant, and x is the displacement of the mass

from equilibrium. For a lattice of N coupled masses, the equations of motion generalize to a system of differential equations. The net force on mass i between any pair of masses mediated using a spring can be described with the equation

$$m_i a_i = - \sum_j k_{i,j} x_j, \quad (5)$$

where i and j are mass indices.

To simulate this system on a quantum computer, the above classical equations of motion must be mapped to the time dependent Schrödinger equation, given by

$$\frac{d}{dt} |\psi(t)\rangle = -iH |\psi(t)\rangle, \quad (6)$$

where $|\psi(t)\rangle$ is the state of the system at a given time and H is the Hamiltonian. As recently shown by Li, an effective mapping is achieved by finding Q such that

$$Q^T Q = \mathbf{k}, \quad (7)$$

where $\mathbf{k}$ is a matrix representing all spring coefficients, and Q^T is the transposed Q matrix [6]. The state of the system is given by

$$|\psi\rangle = \frac{1}{\sqrt{2\mathbf{E}}} \begin{bmatrix} \sqrt{M}\mathbf{v} \\ iQ\mathbf{x} \end{bmatrix}, \quad (8)$$

where $\mathbf{E}$ describes the total energy of the system, $\mathbf{v}$ is the velocity of the masses, and M is a diagonal mass matrix. The Hamiltonian is therefore given by

$$H = - \begin{bmatrix} 0 & \sqrt{M}Q^T \\ Q\sqrt{M} & 0 \end{bmatrix}. \quad (9)$$

Finding the matrix Q is nontrivial, however, Li gives a matrix describing a one-dimensional system with nearest-neighbor and next-nearest neighbor interactions of the form

$$Q = \begin{pmatrix} q_0 & 0 & 0 & \dots & 0 \\ q_1 & q_0 & 0 & \dots & 0 \\ q_2 & q_1 & q_0 & \dots & 0 \\ 0 & q_2 & q_1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & q_0 \end{pmatrix}, \quad (10)$$

where in a mass-spring system, q_0 describes the self restoring force for a given mass, q_1 describes the nearest-neighbor interactions, and q_2 describes the next-nearest-neighbor interactions. Each diagonal band of q_0 , q_1 , and q_2 can, in principle, contain distinct values for each entry, enabling the modeling of inhomogeneous systems.

4. PLUCKED STRING

In the following, I apply this framework to a plucked string modeled as a one-dimensional lattice of N masses with nearest and next-nearest neighbor couplings. To simulate a pluck, the state of the system is initialized such that the masses are displaced in the shape of a triangle; one mass is maximally displaced, and other masses are linearly aligned towards the fixed endpoints. I define the location of the

pluck (i.e. the mass with maximal initial displacement) as x_0 .

The simplest simulation case is one where the endpoints of the string are not fixed, as depicted in Figure 1(b). The result is that energy never leaves the system, and the recovered sound is inharmonic due to the lack of energy dissipation.

To more realistically simulate a plucked string, the endpoints must be fixed, as depicted in Figure 1(a). Because there is no way to evolve different masses under multiple distinct equations of motion using this framework, the string must be simulated as un-fixed, then manually adjusted at each timestep. Manually altering the endpoints to ensure their velocity is always zero therefore removes energy from the string, creating an envelope over the audio waveform. Such a constraint requires that the state of the system is re-normalized after the calculation at each timestep, adding significant computational overhead. The consequence of continually renormalizing the system, however, is that the entire statevector of the system must be precisely known for every timestep, requiring a large number of measurements. A way to work around this issue is by adding damping to the system and ensuring that the endpoint masses have very high damping coefficients. This procedure is beyond the scope of this paper, but it is discussed in the Future Work section.

The velocity of each mass is averaged to produce a single sample at a given timestamp. **Figure 2** shows the velocity of each mass in a sixteen mass system at $t = 0$ when the string is released from the plucked position, as well as snapshots shortly after the release. The string is plucked at the 5th mass. The non-uniform velocity represented at timestamps after the pluck result from assigning each mass a slightly different weight, modeling an imperfect string.

To add another level of realism to the simulation, the string should be additionally coupled to dissipative elements such as the body of an instrument or air surrounding the instrument. If sufficient dissipative elements were to be simulated along with the vibrating string, the energy from the initial string pluck would dissipate throughout the rest of the system, allowing the string vibration amplitude to decay. This requires a new mapping from the classical equation of motion for a damped mass-spring system to the Schrödinger equation, a task beyond the scope of this paper.

4.1 Simulation Procedure

The procedure to simulate a one-dimensional vibrating string without dissipation is as follows:

1. Define the static parameters of the system. This includes the number of masses N , the mass values m_i , the tension coefficients γ , the time step (ex. $1/44,100$), and the total desired audio sample size.
2. Define the initial displacement of the string. In a physical sense, define how the string is pulled from equilibrium just before it is plucked. As an example, the maximally displaced mass x_0 can be displaced by an amplitude $A = 1$. In practice, the masses can be initialized to any collection of displacements.
3. Construct q_0 , q_1 , and q_2 , corresponding to the self restoring force, the interaction strength between a mass

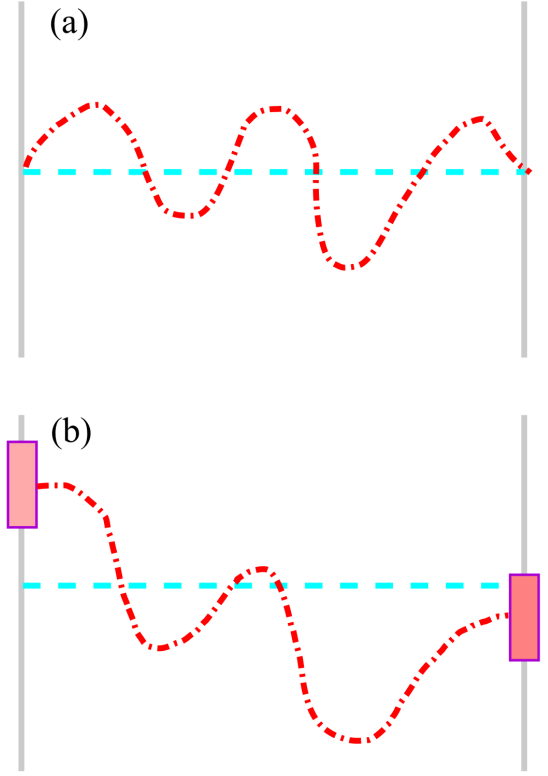


Figure 1. Sketch of a plucked string with fixed (a) and unfixed (b) endpoints. In the fixed case (a), the masses at either extreme of the string are assigned zero velocity. In the unfixed case (b), there is no such restriction. In both the fixed and unfixed cases, each mass in the mass-spring lattice can be thought of as only having freedom of motion along the y-axis. In an experimental setup, an unfixed vibrating string can be approximated by fixing the endpoints of the string to frictionless carts allowed to move along the y-axis as represented by red boxes in (b).

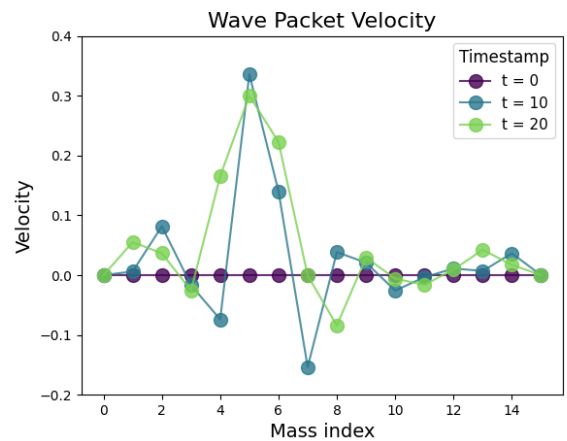


Figure 2. Plot of mass velocity in a sixteen mass mass-spring system with fixed endpoints. The system begins at rest when $t = 0$. Mass number 5 is released, causing it to have great positive velocity at $t = 10$, and the energy wave disperses slightly in either direction by $t = 20$.

and its nearest neighbor, and a mass and its next nearest neighbor. For example,

$$q_0 = \gamma \frac{1 + 1/\sqrt{3}}{2}, \quad (11)$$

$$q_1 = -\gamma, \quad (12)$$

$$\text{and } q_2 = \gamma \frac{1 - 1/\sqrt{3}}{2}, \quad (13)$$

where γ is the tension coefficient defined in the previous step, which can be used to change the resonant frequencies of the string. The coefficient γ can be thought of as a tuning peg on a violin.

4. Construct the matrices Q , H , and the initial input state $|\psi(0)\rangle$.

5. Compile the quantum time-evolution circuit using the input state and Hamiltonian operator. Note that the compilation of the Hamiltonian into a manageable number of qubit logic gates is an active area of research, and the subtleties of this process are beyond the scope of this paper.

6. Execute the quantum circuit and measure the output state.

7. Repeat steps 5 and 6 enough times to adequately reconstruct the state of the system. The number of measurements required for the desired level of accuracy scales linearly with the number of masses. The resulting distribution represents the velocity of each mass at a given timestamp.

8. Sum and normalize the measured velocities to find the net amplitude and save the value as an audio sample.

9. Repeat steps 5 through 8, plugging in a new time value in the time-evolution operator for each sample.

4.2 Fixed Endpoints

If the ends of the string are fixed such that the velocity of the first and N th masses are zero, the quantum state must be renormalized after each timestep. The updated state then becomes the input for the next timestep. This has significant ramifications in terms of measurement complexity. For an unfixed string, only the average velocity of a mass-spring system is required to construct an audio sample, requiring relatively few measurements. In contrast, simulating a fixed-endpoint string requires reconstructing the wavefunction to very high resolution so as not to distort the physics at every timestep. The number of measurements needed to fully reconstruct the wavefunction scales linearly with the number of masses in the system, negating much of the quantum computational advantage. Advantage can be recovered using an amplitude estimation algorithm or with careful application of amplitude amplification as detailed in Section 5.

One additional step inserted after step 7 is required to calculate the dynamics of a one-dimensional string with fixed endpoints: Set the velocity of the first and last mass to zero, and renormalize the state vector. The renormalized state vector is the new input state for step 5.

4.3 Simulation Results

The following analysis focuses on the fixed endpoint plucked string as it produces the harmonic spectrum and amplitude

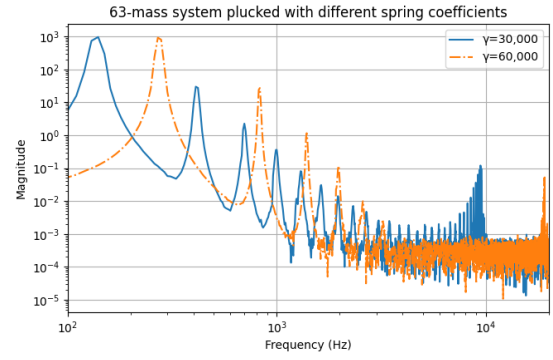


Figure 3. Plot of frequency spectra for mass-spring systems with 63 masses and fixed endpoints 0.05 seconds after the pluck for two different tension coefficients γ . Increasing γ increases the pitch of the simulated string.

decay characteristics of a real string.¹ The most significant difference between the fixed and unfixed cases is that energy never leaves the unfixed string. Initial conditions such as pluck location, irregular masses, etc. result in a complex inharmonic spectra that never decay. This suggests that the unfixed string may be a new and idiomatic form of quantum tone synthesis.

The following plots and discussion are intended to show that the alterable parameters in a mass-spring system are useful in simulating subtle aspects of string vibration in an intuitive way. My approach to varying parameters and exploring the characteristics of the system is guided by my own background as a guitarist. Just as I twist the tuning peg to raise or lower the tension of a guitar string, I can vary the tension parameter γ to shift the frequency spectrum of the plucked string. **Figure 3** shows the frequency spectra of a string with 63 masses for different tension coefficients γ . As the tension grows, so does the fundamental frequency and the harmonic series above it. Due to the small number of masses used to simulate the string, there is an increase in high frequency harmonics approaching 10^4 Hz with γ set to 30,000 which decays after letting the string ring for a short duration.

As a guitarist, I also think about the location from which I pluck the string and how that may affect the brightness of the sound. Plucking closer to where the string is fixed to the bridge of the guitar *sul ponticello* is generally much brighter due to an increased volume of the upper partials compared to a *sul tasto* pluck near the fingerboard of the guitar where upper partials are suppressed. **Figure 4** shows a plucked string plucked from a *ponticello* position at mass 62 next to the fixed endpoint and another string plucked at its midpoint. The resulting spectra show a relative boost in the amplitude of the partials of the *ponticello* pluck near the bridge compared to the *tasto* pluck.

Though I have little control over this in real time as a performer, I have certainly played a guitar strung with old and damaged strings. To emulate the effect of damaging a

¹ The code used to simulate both fixed and unfixed strings using a quantum algorithm, alongside several hundred audio samples, can be found at https://github.com/Xavman42/TENOR26_Mass-Spring_lattice_plucked_string.

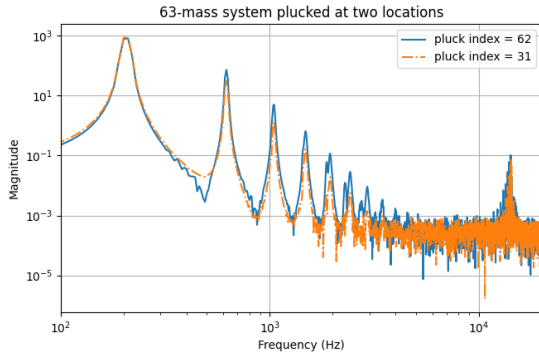


Figure 4. Plot of frequency spectra for mass-spring systems with 63 masses and fixed endpoints 0.05 seconds after the pluck for two different pluck locations. Plucking at the 62nd mass is akin to plucking a string close to the bridge (i.e. *sul ponticello*), whereas plucking at the 31st mass is akin to plucking from the middle of the string. Note how the upper partials are slightly louder in the *ponticello* pluck than the mid-point pluck.

string, I can vary each mass in the system to make it non-uniform. In **Figure 5**, I plot the spectra of a mass-spring system where every mass is identical and compare that with another string where each mass varies by up to 20%. While the fundamental frequency is largely unchanged, some of the upper partials are amplified, and others are attenuated. More strikingly, distinct inharmonic components appear in the attack transient, though they quickly decay after a short time. Perceptually the string sounds more dead, as though the string has been played a lot and for a long time, both damaging the string and adding layers of skin and oil to the surface.

5. VIRTUAL MICROPHONE

The standard method to extract the entire wavefunction is to repeatedly prepare, run, and measure the final state of a quantum circuit to recover $|\psi(t)\rangle$. For the fixed-endpoint simulation described above, this full reconstruction of the state is required at every timestep because the updated state must be precisely known to serve as the input for the next sample calculation. The number of measurements required to reconstruct a distribution up to an arbitrary error scales linearly as the number of masses increases. Considering that the number of masses which can be simulated is given by 2^{n-1} where n is the number of qubits in the system, it is worth investigating if there is some alternative way to extract information from the system.

To limit the number of measurements required to recover useful information from the system, I propose a “virtual microphone” technique based on quantum amplitude amplification. The approach is inspired by how membrane microphones deflect in response to local air pressure variations and convert mechanical energy into an electrical signal. In short, amplitude amplification is a method of increasing the likelihood of measuring one or more states at the end of a quantum calculation. For example, an arbitrary

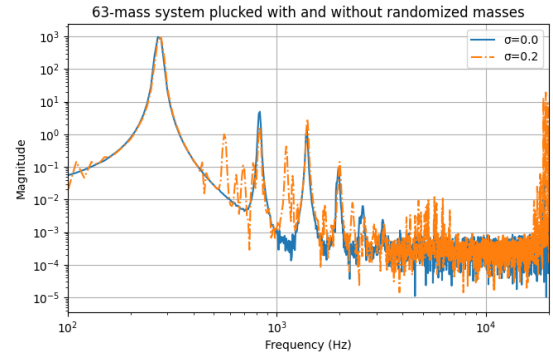


Figure 5. Plot of frequency spectra for mass-spring systems with 63 masses and fixed endpoints 0.05 seconds after the pluck for two different mass distributions. $\sigma = 0.0$ represents a system where every mass is identical, while $\sigma = 0.2$ represents a system where masses vary in size from 80% to 120% of the default mass. Note that additional inharmonic tones are present in the spectra when the masses are not uniform.

encoded state is shown in **Figure 6(a)**, and the state after amplitude amplification is shown in **Figure 6(b)**. The relative magnitude between the target states is preserved, and the likelihood of measuring any other state is suppressed.

Acoustically, humans do not perceive sound as the velocity of every atom in the vicinity. Rather, we detect local pressure variations using the tympanic membrane in the ear. Similarly, a guitar pickup can be thought of as an electromagnetic microphone which measures only the nearest portion of a vibrating string. An idealized guitar pickup for the plucked string described in this paper could be modeled as the velocity over time of only a single mass in the string.

Extracting the specific amplitude of a single mass via amplitude amplification requires defining the target state relative to some reference. The solution I propose is to simulate two identical but classically separated mass-spring systems within the same quantum circuit which differ only by an overall sign in their initial displacement. Both the ordinary and inverted systems evolve identically in time but with opposite velocities for every mass. The joint quantum state of both strings maintain anti-symmetry throughout for any timestamp because they obey the same physics.

Doubling the computational size of the system and accounting for the inverted string can be done at the cost of only a single additional qubit. Amplitude amplification is then applied to a pair of basis states, one mass in the primary system and the other corresponding mass in the inverted reference system. After amplification, the probability ratio of measuring these two states is equal to the square of the target mass’s velocity. Regardless of the number of masses N , estimating a probability ratio between two states requires a number of measurements dependent only on the desired precision. This technique can also be generalized to amplify more than a single pair of states, effectively emulating a microphone with a wider diaphragm.

The amplitude amplification algorithm would be appended at the end of the quantum circuit generated in step 5 in

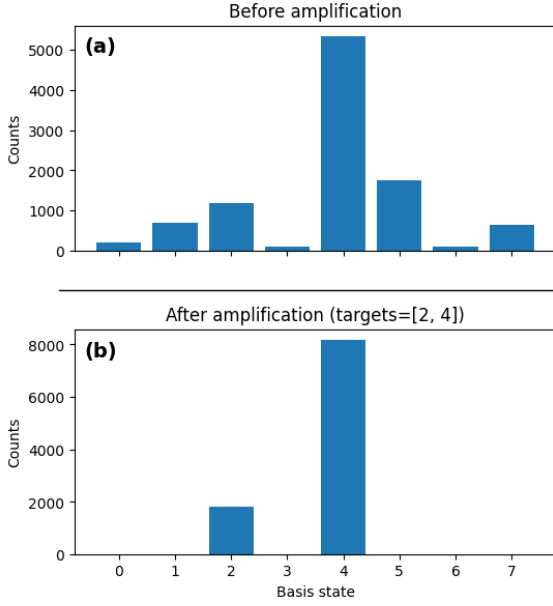


Figure 6. Plots showing the before (a) and after (b) of an amplitude amplification procedure targeting basis states 2 and 4. Note the change of scale in the y-axis. In both cases, the system was measured 10,000 times.

the previous section. Simulating this quantum amplitude amplification is computationally prohibitive for mass-spring systems large enough to be sonically interesting. To validate the sonic implications of this virtual microphone idea, I can instead extract the velocity information of a single mass from the state vector simulations in the previous section.

The frequency characteristics of a plucked string for two virtual microphone locations in a fixed-endpoint system are plotted in **Figure 7**. Measuring a mass at the center of the string generates a waveform exceptionally similar to the average mass velocity from the total statevector measurement. Measuring a mass close to the fixed endpoint (i.e. close to the bridge on a guitar) brings out additional overtones which decay quickly after the initial transient response right after the string is released.

6. DISCUSSION

Mass-spring systems offer an intuitive method for audio physical modeling, but scale exceptionally poorly using any known digital computing algorithm. This paper points out that quantum Hamiltonian simulation can dramatically increase the calculation speed for this system and efficiently model mass-spring lattice at classically intractable scales. I have also proposed a “virtual microphone” technique based on amplitude amplification for efficiently extracting localized audio signals from a quantum state which ensures the same number of system measurements are required to extract useful audio information for any size of mass-spring system.

It is important to emphasize that this quantum simulation framework for mass-spring systems is scalable to any num-

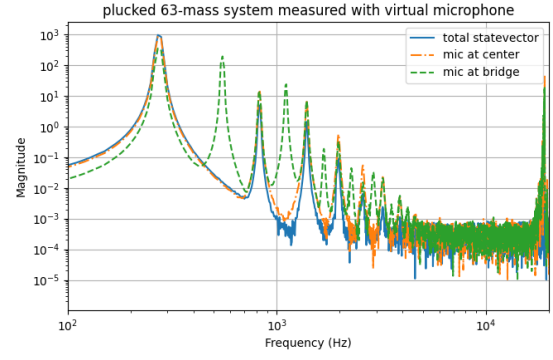


Figure 7. Plot of frequency spectra for mass-spring systems with 63 masses and fixed endpoints 0.05 seconds after the pluck for three measurement methods. The total statevector is the measurement method used in all previous graphs where the velocity of all masses in the system are summed and normalized to create an audio sample. The “mic at bridge” spectrum was created from measurements of just a single mass, close to the end of the string. The “mic at center” spectrum was similarly created using a mass at the center of the string. Note how measuring a mass close to the center of the string provides a reasonable approximation of the entire state vector. Measuring a mass closer to the fixed endpoint instead captures more partials, producing a brighter sound on the attack akin to the bridge pickup on an electric guitar.

ber of spatial dimensions. While circuit depth grows in a multiplicative way as the number of spatial dimensions increases, the exponential scaling in the number of simulated masses remains, [6] meaning that modeling a large-scale three dimensional mass-spring system is computationally feasible. Current noisy intermediate-scale quantum processors host over 100 qubits and can execute thousands of gates in a single circuit, though their noise limits output fidelity [10]. Rapid progression towards fault-tolerant quantum computation suggests that it is not unreasonable to expect the acoustic simulations outlined in this paper to be accessible on actual quantum hardware within a few years.

With a mere 100 qubits, on the order of 10^{29} masses in a mass-spring lattice system could be simulated using the procedure outlined in this paper, in any number of spatial dimensions. The primary caveat to this statement is that only so many operations can be made upon modern qubits before they become subject to noise and decoherence, destroying the data, but hardware is rapidly improving and hardware may one day no longer be a significant restriction. To demonstrate just how large that number is, consider a trumpet. A trumpet weighs approximately 1.2 kilograms, and is made mostly of brass. Yellow brass has the chemical formula CuZn33 with a molar mass of 2.221 kilograms, so a trumpet contains approximately 0.54 moles of brass, equivalent to roughly 3.24×10^{23} atoms. If I also model the air surrounding the brass as a system of weakly coupled masses and springs, then I could estimate that the total system of a trumpet and the air around it could be simulated using approximately 10^{24} masses. In other words, a 100 qubit

computer with sufficient fidelity and gate depth should be able to simulate a system of 10 billion trumpets and the air surrounding the instruments at the atomic level.

Physically modeling instruments using this mass-spring lattice approach requires knowledge about the material makeup of objects. Even the most pure brass alloys have some imperfections which could be reasonably modeled using this method. Simulated wood and other organic materials used to make instruments would also resonate in a significantly more realistic fashion if material irregularities could be simulated. For this to be a realistic synthesis technique, much research is needed around recreating the resonant characteristics of different materials using masses and spring coefficients.

Perhaps the most significant step towards increasing the realism and practicality of this procedure would be the inclusion of damping coefficients in the mass-spring system [11]. Modeling velocity dependent dissipative forces such as viscous friction would naturally produce a sound envelope in the waveform. Further research is required to find the correct mapping from the classical damped mass-spring system to the Schrödinger equation following Li [6]. If damping can be accounted for in the unitary time evolution of the quantum circuit, then there would be no need to renormalize the state vector describing a string fixed at both ends because a given mass can be effectively fixed by over-damping its interactions.

7. CONCLUSION

I have outlined how to simulate the evolution of a one dimensional mass-spring system after it was plucked like a guitar string using a quantum computer. I have also proposed a method of effectively extracting information from the system using a “virtual microphone” technique by duplicating the classical system and applying amplitude amplification to the target masses.

The significance of this procedure lies not in its immediate practicality, rather it lies in its promise to afford us a new way to simulate large and complicated acoustic systems in a precise way without the need of approximations and assumptions made in wave-guide and other synthesis techniques. To accurately simulate sound traveling through materials and transitioning between objects and the air requires understanding how to accurately represent the acoustical behavior of the materials using mass-spring lattices. As the theoretical foundation of this synthesis technique is made more robust and as quantum hardware becomes more accessible, mass-spring systems may prove useful both for the synthesis of realistic sounds and for improving our understanding of how sound propagates through the world around us.

8. REFERENCES

- [1] J. O. Smith, “Physical modeling using digital waveguides,” *Computer Music Journal*, vol. 16, no. 4, pp. 74–91, 1992. [Online]. Available: <http://www.jstor.org/stable/3680470>
- [2] J. M. Adrien, *Representations of Musical Signals*. MIT Press, 1991, ch. The missing link: Modal synthesis, p. 269–297.
- [3] C. Bruyns, “Modal synthesis for arbitrarily shaped objects,” *Computer Music Journal*, vol. 30, no. 3, pp. 22–37, 2006. [Online]. Available: <http://www.jstor.org/stable/4617941>
- [4] K. Karplus and A. Strong, “Digital synthesis of plucked-string and drum timbres,” *Computer Music Journal*, vol. 7, no. 2, pp. 43–55, 1983. [Online]. Available: <http://www.jstor.org/stable/3680062>
- [5] J. Villeneuve and J. Leonard, “Mass-interaction physical models for sound ans multi-sensory creation : Starting anew,” 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:204780346>
- [6] X. Li, “Exponential quantum speedup for simulating classical lattice dynamics,” *Phys. Rev. Lett.*, vol. 135, p. 080602, Aug 2025. [Online]. Available: <https://link.aps.org/doi/10.1103/z2jq-1rxp>
- [7] X. Davenport, “Single sample per qubit audio signal processing: A crash course in quantum computing for computer musicians,” *Emille*, vol. 23, pp. 31–44, 2025.
- [8] C. Bernhardt, *Quantum computing for everyone*. Cambridge: The MIT Press, 2019 - 2019.
- [9] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, 10th ed. Cambridge :: Cambridge University Press, 2000. [Online]. Available: <http://catdir.loc.gov/catdir/samples/cam031/00028911.html>
- [10] “Ibm quantum computing — hardware and roadmap,” <https://www.ibm.com/quantum/hardware>, 2026, accessed: 2026-01-04.
- [11] D. Morgan and S. Qiao, “Analysis of damped mass-spring systems for sound synthesis,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2009, no. 1, p. 947823, Jun 2009. [Online]. Available: <https://doi.org/10.1155/2009/947823>

CRICKET TOOLS: REAL-TIME AUDIO SCORING OF QUANTUM STATE EVOLUTION ACROSS NETWORKED MOBILE DEVICES

Pedro González-Fernández
Escola Superior de Música
de Catalunya, Barcelona
pgonzalez2@esmuc.cat

Thomas Noll
Escola Superior de Música
de Catalunya, Barcelona
tnoll@esmuc.cat

Reiko Yamada
Escola Superior de Música
de Catalunya /
ICFO - Institute of Photonic
Science, Barcelona
ryamada@esmuc.cat

ABSTRACT

This paper presents Cricket Tools, a browser based client-server system for distributed, real-time audio scoring and synchronized audio playback across mobile devices (phones, tablets) connected to the same local network. Cricket Tools supports two complementary approaches: (1) synchronized fixed-media playback (e.g., multi-part stems) and (2) parameterized, real-time audio scores in which each performer’s device synthesizes an individualized audio guide. This second paradigm is designed for situations where the “score” is not a pre-rendered audio or video file but an evolving process controlled by a conductor or an interactive generative system.

In the paper, we position Cricket Tools relative to prior work on audio and audiovisual scoring and distributed notation, drawing on Bhagwati’s framework of audio scores and on Schimana’s “sound as score” practice, as well as on Bell’s SmartVox/SmartBox lineage of web-based distributed parts for ensemble rehearsal and performance.

We also describe the system’s architecture and timing approach, including a lightweight time-synchronization model and client-side scheduling for consistent playback across devices. As a use case, we outline a prototype in which features derived from quantum wavefunction evolution (magnitude/phase descriptors) are mapped to real-time audiovisual scoring parameters distributed to multiple players.

1. INTRODUCTION

1.1 Context and Motivation

Digital notation and time-based score media have broadened the idea of what a “score” can be, especially in screen-based cases where time, motion and interaction become part of the notation itself. In animated notation, we follow Fischer’s understanding of the score as an invitation to a *mapping process* in which performers connect visual attributes to sonic or performative parameters, often using a screen-based temporal reference such as a moving playhead [1].

Within this broader shift, the sound itself has also been treated as a notational medium. Schimana explicitly frames “sound as score” as a way to communicate with musicians through listening rather than written notation, including approaches such as “live generated audio score” (performers imitate sounds they hear) and acoustic memory-based scoring methods [2].

In a related way, Bhagwati proposes “audio scores” as an interface for musical conveyance through sound and develops a vocabulary for how such scores operate in practice and rehearsal. He proposes a set of conveyance modes: information, instruction, imitation, inspiration, and instance; and discusses how indexical instructions can frame other modes. Alongside these classification, Bhagwati notes that audio scores can be implemented as live-generated material or as fixed, repeatable individual tracks, and he identifies synchronization as a remaining technical uncertainty in more “data network-centric” situations with many networked devices [3].



Figure 1. Local-network setup with conductor and multiple player devices connected to a central server.

1.2 Distributed and Networked Scoring on Mobile Devices

A central motivation that follows from these approaches to notation is practical: once a score becomes time-based (audio, audiovisual, animated, or otherwise dynamic), performers need a shared temporal reference while still being

able to work with individualized parts. In work on networked digital scores, Eldridge et al. highlight synchronization between conductor, performers, and notation as a core design issue, alongside performer autonomy and usability in rehearsal [4].

Bell’s SmartVox/SmartBox lineage demonstrates a web-based approach to distributing synchronized parts to phones and tablets using separate performer and conductor clients, where the conductor can start/stop playback and jump to labeled sections; Bell also documents earlier approaches such as iVideoShow (a native iOS app remotely controlled via OSC over Wi-Fi) [5] [6].

More broadly, a substantial body of related work has developed distributed systems for screen-centered score delivery and dynamic visual rendering over local networks, including InScore[7], Quintet.net [8], NetCanvas[9], and ZScore[10], as well as MaxScore[11] and Bach-based[12] workflows, Decibel ScorePlayer[13], dfscore[14], and integrated environments such as Antescofo[15]; middleware-oriented frameworks such as Odot[16] (including OpenMusic integration) or Landini. Our system is situated within this same distributed scoring problem space, but is designed as an audio-score-centered tool.

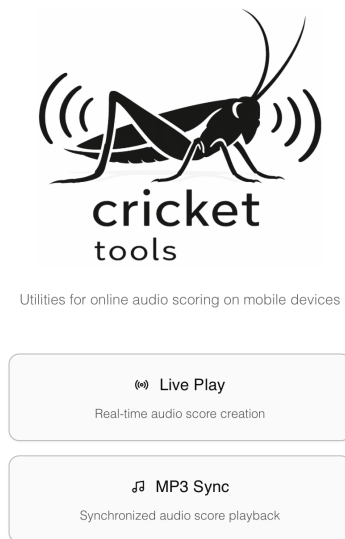


Figure 2. Cricket Tools landing page.

1.3 Cricket Tools

Cricket Tools is a browser-based client–server system for local-network performances where several participants (performers, or audience members) use their own mobile devices as sound score receivers. Its design follows two established needs already present in the literature: (1) reliable synchronized playback of distinct fixed-media parts (as in SmartVox’s distributed parts and earlier mobile playback approaches iVideoShow), and (2) support for “live-generated” or process-driven audio scores where the notational content is produced during performance rather than pre-rendered.

The present paper focuses on the design and technical implementation of Cricket Tools and presents, as an illustra-

tive use case, a prototype mapping of quantum wavefunction evolution descriptors (e.g., magnitude/phase-derived features) into real-time scoring parameters distributed to multiple players, reflecting the original motivation of the project as a distributed quantum sonification tool [17]. The concluding section discusses prospective extensions of the system toward more complex parameterized scores, including microtonal notation and haptics.

2. IMPLEMENTATION

2.1 Overview

The system is organized around two user roles: Conductor and Player, and two tool modes. The first mode (Synth Tool) distributes a parameterized, real-time audio score in which each Player device synthesizes an individualized track accompanied by a simple animated visualization. The second mode (MP3 Sync) distributes pre-rendered fixed media parts to Players and triggers synchronized playback across devices. Both modes share the same networking layer (Socket.IO).

2.2 Software Architecture

Cricket Tools implements a client–server architecture with a Node.js backend and a browser-based frontend. On the server side, Express provides HTTP routing, while Socket.IO provides persistent bidirectional WebSocket communication for real-time state synchronization. The client is a React single-page application bundled with Vite, served by the same Express instance. All conductor dashboards and player interfaces connect to the server through Socket.IO and receive state updates as events rather than through polling. Fixed-media operations (upload/download/delete) are implemented as HTTP endpoints, while real-time performance controls and state changes propagate through Socket.IO events. Server state is held in memory (players, MP3 slots, and current playback state), which simplifies lookup and broadcast but implies that state resets on server restart.

The implementation uses a simple clock synchronization mechanism to estimate each client’s offset from server time. The client sends a `syncPing` containing a local timestamp; the server replies with `syncPong` including the original client time and its own server timestamp. The client measures round-trip time and estimates the server time at reception assuming approximately symmetric network delays, then stores the resulting `serverTimeOffset`. After this step, clients estimate the “current server time” as `Date.now() + serverTimeOffset`, and use this estimate when scheduling audio events. This offset is used in both tools: in the Synth Tool it supports phase-aligned cyclic triggering; in MP3 Sync it enables scheduling playback against a future server start time with a safety buffer.

2.3 Synth Tool: Distributed Real-Time Audio Scoring

In the Synth Tool, Players join through a browser interface by entering a display name; the server assigns each Player a stable numeric identifier (1, 2, 3, ...) for the duration of the session. This provides a consistent address-

ing scheme for both the conductor dashboard and external control (Section 2.4). Each Player's state includes at minimum: playerId, socketId, name, pitch (MIDI), interval (cycle duration in ms), and a phaseStartServerTime used for phase alignment.

The conductor dashboard displays all connected Players and provides real-time controls for at least two parameters per Player: pitch (e.g., MIDI 36–84) and interval (e.g., 50–3000 ms). When the conductor changes a control, the client emits a Socket.IO message to the server. The server updates the in-memory Player state and broadcasts (i) a targeted update to the affected Player and (ii) a global state update to all conductor clients. This design ensures that the conductor interfaces remain consistent.

On the Player side, audio is generated locally using Tone.js. Each Player instantiates a synthesizer and produces discrete note events based on the current phase of a cyclic timeline. The phase is computed from estimated server time and the server-provided phaseStartServerTime and interval, so that Players sharing the same phase reference can align their trigger moments even if they joined at different times. In the current prototype, Players also render a circular visual representation of the cycle, supporting the interpretation of the audio guide.

A key implementation detail concerns continuity when the conductor changes tempo/interval. Naively replacing the interval would shift phase and produce discontinuities in timing. To preserve continuity, the server computes the Player's current phase under the old interval and adjusts phaseStartServerTime so that the Player continues from the same phase position under the new interval. This allows performers to experience parameter changes as continuous transformations rather than abrupt resets, while keeping the phase model consistent across conductor and Player clients.

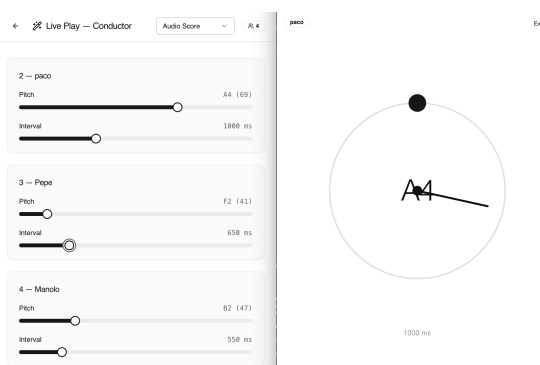


Figure 3. Synth Tool interface showing conductor controls and player view for real-time audio scoring.

2.4 Control from Max/MSP: Node-For-Max Bridge

A core design goal is to allow algorithmic control from composition environments without requiring manual interaction with the conductor dashboard. Cricket Tools provides a Node for Max module (conductor-control.js) that connects to the server via Socket.IO and registers itself as a conductor client on connection. Once connected, the script

accepts Max messages and translates them into Socket.IO control events (maxCommand) that follow a compact integer format: target, control, value. The target selects a specific Player by numeric playerId or broadcasts to all Players (e.g., target = -1). The control selects a parameter (e.g., pitch or interval), and value provides the new parameter value.

On the server, incoming maxCommand events are validated as conductor-originated, then routed through a control resolver that maps control IDs to named parameters and applies the update to the addressed Player(s). The server maintains a bidirectional mapping between numeric player IDs and socket IDs, so that a Max patch can address Players consistently even as their socket connections change. The bridge also supports bidirectional feedback by receiving server state updates and outputting basic state information to Max (e.g., the number of connected Players), enabling patches to adapt to the number of players in the ensemble. An OSC interface can also be supported on the server side via UDP, but in the current implementation the Node-for-Max WebSocket bridge is the primary external control pathway.

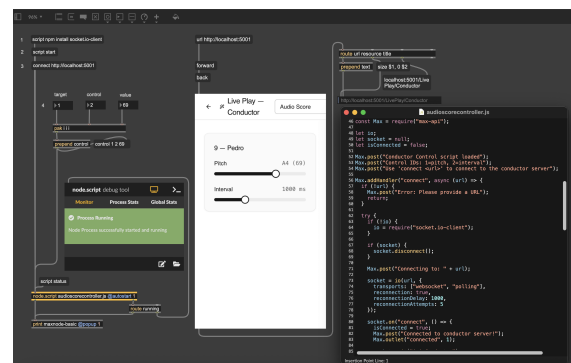


Figure 4. Node-for-Max WebSocket interface enabling algorithmic control of the Synth Tool from Max/MSP.

2.5 Mp3 Sync Tool

The MP3 Sync Tool distributes per-part MP3 files and triggers synchronized playback. It uses a fixed slot model (eight slots), where each slot can hold one player and one assigned file; file assignment persists across player disconnects. The conductor uploads files via HTTP and Players download and decode them locally. For playback, the server broadcasts a future start time and optional seek position; Players schedule Web Audio playback using their estimated server-time offset. Late joiners can be instructed to start from the current playhead derived from elapsed time since the server start.

3. USE CASE: “COLLECTIVE PERFORMANCE OF QUANTUM WAVE FUNCTIONS”

In the present section we outline a specific and inventive scenario of an explorative musical performance, which seeks a particular interaction between a social and a mathematical concept of unity. In classical chamber music (like a

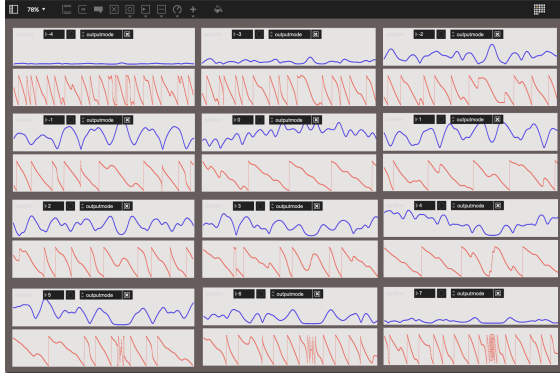


Figure 5. Max/MSP patch visualizing quantum wavefunction evolution (magnitude and phase) and mapping it to real-time audio-score control data.

string quartet), the goal is often the homogenization of sound—the four instruments striving to sound like one single, 16-stringed instrument. Our quantum model takes this “social unity” and provides it with a mathematical backbone: the Non-locality of a quantum wave function.

What we call the *use case* here, is actually first and conceptually a reduced experiment, which serves as an intermediate goal for a more advanced project. It is nevertheless worthwhile to sketch the overall idea, not least thanks to prominent logistic role of the Cricket Tools.

3.1 What does it mean to perform a one-dimensional wave function?

Imagine n musicians sitting in a circle. Each musician represents a single point in space (a position x in the set $\mathbb{Z}_n$). The “one-dimensional” aspect simply means we are looking at a line of data that wraps around—like a string that starts at Musician 1 and ends back at Musician n . At each moment the time-development of the wave function provides each musician with a complex number $\Psi(x, t) = ae^{ip}$, which can be represented by a magnitude $a \geq 0$ and a phase $-\pi \leq p < \pi$. Together these values represent a global unity which is expressed by the Hamilton operator of an underlying quantum system.

In “old-fashioned” chamber music, unity is achieved through a collective interpretation of a score, through empathic listening to each other, through gestures and eye contact. From this tradition we inherit a kind “Non-Soloist” Ethos for the present project: Just as in a Haydn quartet where the inner voices (viola/second violin) are vital for the texture even when they do not have the melody, the musicians at positions x where $|\Psi(x, t)|^2$ is low are still essential. They carry the phase, which is the “potential” for the next move. A shared wavefunction acts as a mathematical version of the “breath” that chamber musicians use to start a phrase together and which lets the ensemble act as a single organism. In an earlier contribution ([17] we developed ideas and tools for the musical sonification of quantum wave functions with the means of computer generated sounds. In our Max/MSP package *Qte* one may specify a Hamiltonian H of a (disturbed) Harmonic oscillator, calculate the eigenstates $\phi_n(x)$, choose an initial state ψ and store

its time development $\Psi(x, t)$ as a parameter family for the combined control of a family of tracks in a sequencer[18].

With the present approach we meet the following challenge: How can we inherit the spirit of the social interaction in an ensemble performance on the basis of a fully determined time development — where at first sight the musicians would be degraded to be oscillators for the computer-generated flow of data? The answer is obvious: The musicians should not only share the wave function but they should internalize the entire quantum system from the Hamiltonian H , its eigenstates $\phi_n(x)$, the choice of an initial state $\psi(x)$ in order to know and experience the wavefunction $\Psi(x, t)$ as their collective creation. In combination with the *Qte*-package the *Cricket Tools* will help the musicians to immerse themselves as an ensemble into this process of (self)-discovery.

3.2 The Hamiltonian as the Rulebook of Relationships for the Quantum-Ensemble

The “collective rehearsal” starts with the collective design of a Hamilton-operator H (instead of using a predefined operator from the physics classroom). In the position basis, the Hamiltonian H is an $n \times n$ matrix. Its coefficients can be divided into two categories: the Diagonal and the Off-Diagonal. It is important to distinguish between the meanings of the position and the attitudes of the players at these positions. For example, the real diagonal-coefficient H_{xx} at position x represents the “Potential Energy” at that spot. It is the “Gravity” or “Persistence” of this position. If H_{xx} is high compared to others, that position is “expensive” for the wave to stay in. The wave, i.e. the music, will want to move away quickly. If all H_{xx} are the same, the circle is “flat” and perfectly symmetrical. In other words, the diagonal of H is the “topography” of the stage. High tension acts like an obstacle; low tension acts like a welcoming valley for the music to settle in.

The off-diagonal coefficients represent the interaction dispositions of the ensemble. These are complex numbers and high absolute values H_{xy} stand for a higher degree of “openness” between the player at x and the player at y . It determines how fast music is transmitted between the two. The absolute values coincide $|H_{xy}| = |H_{yx}|$ which means that the amount of sending and attentive receiving have to coincide. The actual direction depends on various factors, such as the initial state. But it can also be influenced by the inverse phases of H_{xy} and H_{yx} , through the activation of a phase rotation which induces a direction of the flow on the circle.

For the more advanced project we envisage a time-dependent Hamiltonian, where a small ensemble may change its interaction rules during the performance. For the present use-case we envisage a larger ensemble with a fixed and simplified Hamiltonian[19].

3.3 Our Scenario: The Homogeneous Circle-Ensemble

This scenario investigates the realization of the idea within a large ensemble arranged in a periodic lattice. The goal is to exemplify the movement of a quantum wave packet

through a medium of uniform tension and symmetric coupling. The ensemble of n musicians is mapped onto the cyclic group $\mathbb{Z}_n$. The collective state $|\psi(t)\rangle$ exists in an n -dimensional Hilbert space $\mathcal{H} \cong \mathbb{C}^n$:

$$|\psi(t)\rangle = \begin{pmatrix} c_0(t) \\ c_1(t) \\ \vdots \\ c_{n-1}(t) \end{pmatrix}$$

where each coefficient $c_x(t) = \langle x|\psi(t)\rangle$ provides the specific performance instructions for musician x .

To model a uniform communicative disposition, we employ a nearest-neighbor Hamiltonian. For an ensemble of size n , the matrix H is defined as:

$$H = \begin{pmatrix} E_0 & -ve^{i\phi} & 0 & \dots & -ve^{-i\phi} \\ -ve^{-i\phi} & E_1 & -ve^{i\phi} & \dots & 0 \\ 0 & -ve^{-i\phi} & E_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -ve^{i\phi} & 0 & 0 & \dots & E_{n-1} \end{pmatrix}$$

We interpret the parameters as follows:

- **Diagonal** ($E_0, \dots, E_{n-1}$): Represents the “Tension” of the stage. In the special case $E_0 = \dots = E_{n-1}$ the distribution of potential energy on the stage is homogenous.
- **Off-Diagonal** (v): The coupling strength or “Bridge.” It determines the speed at which the music flows between neighbors.
- **Phase** (ϕ): The “Chiral Current.” Because $H_{j,j+1} = (H_{j+1,j})^*$, the reciprocity is preserved, but a non-zero ϕ breaks time-reversal symmetry, causing the wave to “spin” directionally around the circle.

3.4 Logistics of the communication with the cricket tools

We assume that each player has her/his mobile phone equipped with a set of headphones, and connects via our QR-code to the local network (via `:// <server-ip>: 5001`) and the cricket-tool-interface in their browser. They access the real-time synth tool, they type their name, receive their player number and accept the usage of the audio system in their phones. In this use case only the conductor can access the Max/MSP-Qte package and the Node-for-Max WebSocket bridge with all the specifications of the Hamiltonian, the initial state and the scorification (parameter settings for the time developments and the actual sound scoring).

4. CONCLUSIONS AND FUTURE WORK

In the quantum-wavefunction use case, Cricket Tools was used in a deliberately reduced setup to distribute magnitude and phase derived descriptors as real-time scoring parameters to multiple players. Within this constrained scenario,

the workflow suggested that parameter-stream-based audio scoring can function in a music ensemble setup: performers were generally able to follow the cyclic cueing logic and respond to conductor-side parameter changes with ease. Future work will therefore focus on extending the affordances of the system while preserving interpretability and coherence across devices.

Beyond the current parameter set, Cricket Tools could be extended to support tuning- and intonation-sensitive scoring where pitch is specified not only as a note name but as an explicit mapping to exact frequencies. One concrete direction is to map the same set of twelve note names onto multiple temperament systems, so that the exact frequencies associated with a given pitch class are recalculated whenever the tuning changes. This would turn the distributed audio score into a “tuning-aware” guide in which performers receive not only pitch information but an explicitly defined intonation context, enabling controlled micro-deviations that remain coherent across the ensemble.

Musically, such tuning transformations could be deployed as recurring formal events: short, clearly delimited sections in which the distributed parts shift temperament in a coordinated way, alternating with intervening passages based on pre-written material that remains fixed within a single tuning system. To heighten perceptual contrast and stabilize the listening frame, one performer (or a dedicated electronic reference part) could remain in a constant tuning throughout the piece, functioning as an anchor against which each reappearance of the “variable temperament” sections becomes newly legible. This proposal remains exploratory, but it suggests how the same infrastructure used for real-time parameter distribution can also support compositional strategies based on repeatable, structurally articulated changes in intonation.

From an implementation standpoint, this extension can be treated as a direct generalization of the current Synth Tool parameter model. Alongside per-player pitch and interval, the server would maintain a global (or per-player) tuning state, selecting a temperament preset or a transmitted tuning table that maps the twelve pitch classes to cents offsets (relative to 12-TET) or to absolute frequencies. To preserve musical clarity and ensemble stability, tuning changes should be scheduled rather than applied immediately: the server can quantize a retuning event to a musically meaningful boundary (e.g., the next cycle onset) or broadcast it as a future server-time update, reusing the same timing approach already employed for phase alignment and for continuity under interval changes. In this way, retuning is perceived as a controlled transformation rather than an abrupt reset, while remaining consistent across devices.

The Node-for-Max bridge provides a natural pathway for algorithmic intonation control. Instead of manual conductor intervention, a Max/MSP process can compute and send tuning changes as part of the same control stream that already addresses player parameters: selecting between predefined temperament families, interpolating between tuning tables over a specified duration, or generating custom microtonal offsets based on higher-level musical logic. In

the quantum-wavefunction use case outlined in this paper, this creates a particularly direct expansion of the mapping space: magnitude- and phase-derived descriptors could be used not only to drive timbral or rhythmic parameters, but also to shape an ensemble-wide “intonation field” (e.g., phase features selecting temperament regions or specifying continuous cents deviations), while the fixed-tuning reference part provides a perceptual baseline against which such changes can be heard as form-defining shifts.

Finally, because microtonal retuning is both perceptually subtle and cognitively demanding, future versions could strengthen multimodal cueing around tuning events. Player interfaces could display a concise indicator of the current temperament (name and/or deviation profile) and an anticipatory cue when a tuning change is scheduled (e.g., a countdown ring aligned to the cycle). Where devices allow it, brief haptic cues could reinforce the moment of retuning without requiring constant screen attention. Combined with Cricket Tools’ dual paradigm (live-generated audio guides in the Synth Tool and fixed reference materials in MP3 Sync) these additions point toward hybrid scores in which intonation functions as a repeatable, structurally articulated parameter: precisely timed, algorithmically controllable, perceptually anchored, and distributed reliably across a networked ensemble.

5. REFERENCES

- [1] C. M. Fischer, “Understanding animated notation,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2015*, Paris, France, 2015, pp. 31–38.
- [2] E. Schimana, “Sound as score: Live generated audio score and audio score based on acoustic memory,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2019*, Melbourne, Australia, 2019, pp. 33–37.
- [3] S. Bhagwati, “Elaborate audio scores: Concepts, affordances and tools,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2018*, Montreal, Canada, 2018, pp. 24–32.
- [4] A. Eldridge, E. Hughes, and C. Kiefer, “Designing dynamic networked scores to enhance the experience of ensemble music making,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2016*, Cambridge, UK, 2016, pp. 193–199.
- [5] J. Bell and B. Matuszewski, “Smartvox. a web-based distributed media player as notation tool for choral practices,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2017*, A Coruña, Spain, 2017, pp. 99–104.
- [6] J. Bell, “Audiovisual scores and parts synchronized over the web,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2018*, Montreal, Canada, 2018, pp. 17–23.
- [7] D. Fober, S. Letz, Y. Orlarey, and F. Bevilacqua, “Programming interactive music scores with inscore,” in *Proc. Sound and Music Computing Conf. (SMC 2013)*, Stockholm, Sweden, 2013, pp. 185–190.
- [8] G. Hajdu, “Quintet.net: An environment for composing and performing music on the internet,” *Leonardo Music Journal*, vol. 15, pp. 23–30, 2005.
- [9] B. Carey and G. Hajdu, “Netscore: an image server/client package for transmitting notated music to browser and virtual reality interfaces,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2016*, Cambridge, UK, 2016, pp. 151–156.
- [10] S. Zagorac and P. Alessandrini, “Zscore: A distributed system for integrated mixed music composition and performance,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2018*, Montreal, Canada, 2018, pp. 62–70.
- [11] N. Didkovsky and G. Hajdu, “Maxscore: Music notation in max/msp,” in *Proc. Int. Computer Music Conf. (ICMC 2008)*, 2008.
- [12] A. Agostini and D. Ghisi, “bach: An environment for computer-aided composition in max,” in *Proc. Int. Computer Music Conf. (ICMC 2012)*, Ljubljana, Slovenia, 2012.
- [13] C. Hope and L. Vickery, “The decibel scoreplayer - a digital tool for reading graphic notation,” in *Proc. Int. Conf. on Technologies for Music Notation and Representation - TENOR 2015*, Paris, France, 2015, pp. 58–69.
- [14] R. Constanzo, “dfscore,” Online, 2016, available: <http://www.rodriigoconstanzo.com/dfscore/> (accessed Jan. 7, 2026).
- [15] R. Burloiu, A. Cont, and A. Poncelet, “A visual framework for dynamic mixed music notation,” *J. New Music Research*, vol. 46, no. 4, pp. 321–332, 2017.
- [16] J. MacCallum, R. Gottfried, I. Rostovtsev, J. Bresson, and A. Freed, “Dynamic message-oriented middleware with open sound control and odot,” in *Proc. Int. Computer Music Conf. (ICMC 2015)*, 2015.
- [17] P. González-Fernández, F. P. Olives, E. R. Bertram, T. Noll, and P. beim Graben, “Quantum automation of playful heterophonic processes,” in *Proc. Int. Computer Music Conf. (ICMC 2025)*, Boston, MA, USA, 2025, pp. 460–466.
- [18] T. Noll, F. P. Olives, P. González-Fernández, and P. beim Graben, “Exploring quantum tonality,” in *3rd International Symposium on Quantum Computing and Musical Creativity (ISQCMC '25)*, Palermo, Oct. 2025. [Online]. Available: <https://doi.org/10.5281/zenodo.18203933>
- [19] T. Noll, P. González-Fernández, and R. Yamada, “The quartet: Embodying quantum theory as an ensemble,” 2026, manuscript submitted for publication.

BEYOND NOTATION: LI-YUE: FROM CENTRALIZED SPECTACLE TO A COMMUNITY MODEL

Eric Chi-Wei LIN

Independent artist-researcher, Paris
chiweimail@gmail.com

ABSTRACT

Modern music theory has largely understood musical organization through the realization of musical works. This paper argues that Li-Yue is best understood as a **protocol tradition** rather than a music work tradition. Instead of producing autonomous musical objects, Yue employed sound more as an operational medium for coordinating bodies, generating affects and social relations through embodied procedures.

Drawing on historical ritual manuals, together with Lin Chiwei's participatory Sound Modules, this article develops a protocol-based account of musical organization. In this view, composition no longer consists of specifying sonic forms, but in designing new conditions under which collective structure emerges through interaction. Sound functions less as material for aesthetic object, than as an operational "signal" within a dynamic process of social coordination.

Reconsidering Yue as **protocol tradition** not only offers a new reading of early Chinese Music practice, but also provides a conceptual framework for contemporary music composition. Rather than producing musical works or spectacles, protocol-based sound practices explore how communities may be generated through shared procedures, distributed participation, and collective enactment.

1. INTRODUCTION

Beginning with Montesquieu, Rousseau, Hegel, Western philosophers have often dismissed East Asian art and music—as functional, repetitive, artistically underdeveloped, or “ahistorical”. Over the twentieth century, these criticisms were revised in significant ways. Scholars across musicology, sociology, ethnomusicology, and cultural anthropology began to critique their own academic legacies: challenging the Eurocentrism of

harmony-tonality, questioning formalism and “score-centered” epistemology.

Today, libraries are filled with field reports on Eastern music. Research grounded in Sinology and East Asian studies is increasingly visible, accompanied by extensive academic work in conservatory music and the sociological studies. Indeed, postwar scholarship has already produced generations of critical paradigms.

Yet revisiting 18th-century European philosophers' attitude toward Asian music is far from an obsolete exercise. These early critiques often reveal—more directly than their successors—the deep ontological misalignments between Eastern and Western theories. Put it strait forwardly: What if the idea of music (樂) spoken of in the Eastern classical context (Yue in modern Chinese, Gaku in Japanese, Ak in Korean)—was never meant to align with the Western notion of “Music”?

This foundational misalignment—due in part to the lack of integration between sinology, musicology, and social theory—has persisted across time. Even after full-scale Westernization in East Asia, the gap has not closed. It has not only shaped how the west misreads eastern music, but also how the East, through self-colonizing academic frameworks, scholars in the East inherit and apply Western categories to interpret their own traditions.

1.1 What Is “Yue”?

In classical Chinese, Yue(樂) appears in conjunction with Li (禮), forming the compound Li-Yue (禮樂). Sinologists often translate them as “ritual and music,” but such a translation is rather insufficient—if not outright incorrect. “Li” refers to a codified structure of rituals that presents order; “Yue” is the sensorial and affective interface through which “Li” becomes operable, while “music” plays an essential role within it. In any case, it is not 'Music' as we understand it today, but a cross-sensory, intermedia system that bridges the related actors.

The YueJi (Book of Music), compiled during the Han dynasty (ca. 130 BCE), expresses the Confucian view succinctly: “Yue arises from within and manifests outward.” and the manifestation is being organized

Copyright: © 2026 First author et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

through “Li-Yue” by the orchestration of affective states, bodily rhythms, synchronized gestures, and perceptual ordering.

Dance was at the core of Yue. In pre-Qin antiquity, all music presupposed movement. Dance was not an embellishment, but the primary way rhythm inscribed itself in the body. As the Yueji repeatedly emphasizes: “Where Yue arrives, dance begins.” Movement was not an optional addition—it was the visible, kinetic articulation of the system’s very rhythm.

Likewise, music in the narrow sense—voices, instruments, scores—was not the sole content of Yue. Sound played in the rituals generally has a functional role: to align with Heaven, stir emotion, synchronize rhythm, and enable coordination. Musical components such as melody and rhythm were not ends in themselves. Sound served as a temporal structure of body/spiritual synchronization—a frame for ritual.

1.2 *Li-Yue* as Ritual Topology

The “theatricality” within *Li-Yue* bore little resemblance to Western notions of drama. It was neither mimetic, psychological, nor narrative. There were no characters with inner lives. Instead, its logic was closer to ritual choreography—role allocation, spatial configuration, and procedural sequencing. Its aim was not representation, but the emphasizing of relational position within an ethical order.

Poems, chants, and liturgical texts served as mediators between sound and ethical position—embedding historical memory and guiding affective direction....Court garments, differentiated by rank and role, became wearable diagrams of hierarchical order. Spatial placements, musical instruments, ritual objects, and crafted artifacts together composed a legible visual grammar.

The architecture of Yue was not neutral, but a topological structure that embodied the cosmological symbol of the very ritual. Spaces such as ancestral temples, altars, or imperial courts were sacred sites of encounter between Heaven and human. Their spatial flow and acoustics represent holy/secular, male/female, near/far, intimacy/distance, ruler/subject divisions.

Even offerings—the placement, types, and timing of sacrificial items—materialized the rhythm and logic of cosmological order. They were not only symbolic, but also operational markers of time and space.

1.3 Not Music, Not *Gesamtkunstwerk*: Yue as Protocol

By definition, ancient Yue was neither a collection of musical “works” nor a form of total art (*Gesamtkun-*

stwerk) in the Western sense. Composition, as understood in Western art music, operates through the development of sonic materials toward a finished work—a self-contained musical object. Yue, by contrast, functioned as a device for producing **embodied order** through rhythm, rather than presupposing any musical object, the Li-Yue Model operated through node-based time, supplying control signals that enabled modular execution—such as for dance and ritual respects. In this sense, Yue is employed as a corporeal–normative system for generating ethical order, not so much as a musical work, but as a mechanism for coordination.

Many surviving Yue texts can be accurately understood as manuals—protocols for internalizing feudal cosmology through ritualized embodiment. A good example can be found in “**Notes on Li-Yue in the Confucian School**” (Shengmen Li-Yue Zhi 聖門禮樂誌), a comprehensive manual for the execution of the Confucian Memorial Ritual edited in 1716. The book presents an extraordinarily detailed prescriptive system: meticulous instructions for the building and tuning of each instrument; specifications of finger techniques; precise indications of when particular sonic patterns should occur within the ritual; and exact spatial positions within the temple from which they are to be performed. The level of its technical precision is doubtless remarkable for its time. [1]

Notably, each percussion instrument (totally 16 kinds) assigned with specific rhythmic patterns, is carefully described. By contrast, the melodic dimensions of wind and string instruments are largely left undefined, or reducing to minimal indications spanning barely a single page. Meanwhile, the book includes three full pages description of rhythmic notation for the large drum and half a page for the turtle drum. This asymmetry reveals a distinctive conception of classic orchestration; the ritual ensemble is conceived as a composite of seven types of wind instruments, two types of string instruments, and sixteen types of percussion instruments, while the exact number of performers and instrumental allocations remained adjustable, the full ensemble typically involved approximately eighty musicians.

From a Western perspective, what is most striking is the absence of an **orchestral score**. The fact behind the design is, Such a system would be only conceivable with **Modular Modes** of organization and with embodied training of musicians. Each musician group functioning according to memorized codes and pre-determined melodic patterns. Rather than being under the sole score and coordinated through a conductor, the ensemble operated through distributed knowledge of rule sets. This once again underscores that Li-Yue functioned not as a centralized representational system, but as a set of operational and procedural modules.

In the Confucian Memorial Ritual, bells and stone chimes serve as temporal markers within the ritual sequence. Bells typically announce the beginning of major

ceremonial sections and punctuate significant transitions. Stone chimes, owing to their shorter and more clearly bounded resonance, articulate internal divisions and structural points within the unfolding ritual. Together they do not regulate the performance through continuous command, but provide a shared temporal framework through which music, dance, and ceremonial action remain synchronized. In this sense, sound functions less as musical material than as a medium of collective coordination.

Each Confucian Memorial ceremony consists of a series of independent modules—action units that can be rearranged and redeployed across different sacrificial contexts (Heaven, ancestors, soil, or grain deities), corresponding to different ritual subjects. While the object and scale varied, the underlying structure remains constant. In this sense, the “Yue” works like an ever-evolving interface through historical time, in order to meet very different needs: an embodied mediation through which updated ethical and political codes became accessible, trainable, and internalized by participants.

We can consider the Yue system as a whole thus operated as a **protocol**—not just a set of contents, but a set of conditional rules that governed interaction. The execution of these rules existed only in runtime. When Yue is framed as an artistic object or compositional work, it ossifies and loses its essence. As a protocol, Li-Yue lives only within a living system of coordinated action.

During the historical course of modernization, from dynastic kingdoms to modern state-nation, Yue’s fate is inseparable from the broader transition: Along with the fall of the empire, the very function of Yue was fragmented, redefined, and redistributed to different institutions.

A decisive rupture occurred with the great westernization in Asia. The western “Art Music” occupied the yet young Asian national imagination. Symphony Orchestra and its associated institutions—such as conservatories—which effectively assumed functions once fulfilled by **Wen-Yue** (文樂, *Gagaku* in the Japanese context).

At the same time, the military music **Wu-Yue** (武樂, literally “army music”) was displaced by Prussian-style march music. Classical empire ritual music was largely erased (except in Japan). The remaining ceremonies fundamentally reconfigured, translated into a westernized ceremonial format, aligned with modern state’s culture apparatus in general.

As far as my main concern relies on musical protocol instead of “Classic Chinese Music”, the following chapter will focus exclusively on the transformation of national ceremony and re-examines its radical consequences.

2. CONTEMPORARY RITUALS

What, then, are the contemporary enactments of Yue?

Beyond the limited number of ceremonies preserved through cultural heritagisation—such as those maintained by the Imperial Household Agency in Japan or at the Munmyo Shrine in Korea, which persist largely as heritage artifacts with little vital forces—the ceremonial field is otherwise dominated by western styles. These ceremonies, embedded within compulsory national education systems, military drills, and televised mass propaganda, have remained largely unexamined. It is precisely within these modern ceremonies that the enduring logic of Yue continues to operate, albeit under transforming conditions. Here, Yue may be observed functioning across three major categories:

2.1 Festivals–Spectacular

Both East and West rely on grand ceremonies. But they play different logic. In the West, events like Bastille Day or 2023 Olympics opening are subject to critique. Itself can as well either be satire or humorous. Performer’s bodies are individuated and styles are plural. While the North Korea’s Arirang mass game or China’s 2008 Olympic ceremony center on synchronized mass embodiment. The goal is as affirmation as representation. The Music here functioned as a metronome and the bodies are anonymized into pixels. This is Yue in its most extreme visualized form—proving the perfect execution of order.

2.2 Education–Body Regulation

Flag-raising ceremonies, anthem singing and military drills erase the performer–audience divide, as do Japanese and Chinese radio calisthenics. Every participant becomes part of one disciplinary rhythm. Eastern ceremonies presume: ‘Regulate the body, then align the mind.’ (Book of Li, ca. 130 BCE), and they are often long, repetitive, and physically exhausting. Embodied discipline here outweighs ideological teaching. Japanese calisthenics models demonstrate normative, affective, educative nature, a sustained social rhythm embedded with daily life (with part of its origin in Zen Buddhism practices), with the main focus on internalized norm enacted through the body. Compared to Prussian calisthenics, it lacks the heroic implication and is closer to traditional ritual drill with strong civilian hygiene implication. [2]

2.3 Military–Sovereign

From Qing-era military reviews, to Emperor Meiji’s naval parades in Japan, to formation-based ceremonies in modern China, Japan, Korea, Vietnam, and Taiwan, Wu-Yue (military yue) continues its march. Where Western parades often appeal to republican traditions, civic identity, and public pride, East Asian spectacles more readily project a claim to cosmic legitimacy—order presented as destiny. Music functions here as a metrical

governor: the parade becomes a mode of sonic governance, a residue of imperial ritual logic.

Yet it is difficult to argue that these twentieth-century state rituals are along the Yue heritage. Modern Rituals core operational grammars share a genealogy with nineteenth-century European militarized education and athletic nationalism; this grammar spread across East Asia through Japan's colonial and military apparatuses.

Anyway, crucial institutional differences remain. Modern rituals run through training and selection pipelines operated by the military, schools, sports clubs, and party-state apparatuses, whereas classical Li-Yue relied on ancestral temples and court/ritual institutions—ritual officials, music officials, and ritual manuals. Modern state rituals typically draw their legitimating vocabulary from the nation-state, revolutionary history, and narratives of achievement in modernization, while classical Li-Yue more explicitly grounded in heaven-human order and moral transformation.

It is reasonable to describe these national ceremonies as a structural homology of classical Li-Yue. The homology rest on shared logics of embodied coordination: synchronization, role assignment, spatial-topological installation (routes, grids, formations), regimes of visibility and constant correction, bodily discipline, and staged legitimization. More than this, We can further find some interior connections such as cosmological believes residue persists in modern contexts—not only in South Korea's Confucian rites, but also, more discreetly, in Chiang Kai-shek-era Taiwan, drills, daily flag-raising, and militarized school training, are integrate into Confucius dogmas, where Chiang repeatedly invoked the “Dao-heritage of five thousand years,” a phrase he often recited.

In the twenty-first century, with its secularism and hyper-technologization, Eastern Asian states continue to be shaped by the logics of Li-Yue. Stripped of its pre-Qin feudal foundations, interiority and generative ethical structure, the Li-Yue system has evolved into a technical module of governance and bodily discipline.

Today the national Li-Yue becomes a fantastic spectacle: Yet this spectacle rendering power as visible, verifiable, and closed, as the final output of a state-aesthetic system, while participants bodies are mobilized, emotions intensified, every elements designed and aligned to meet the ideal of state-nation.

However, If we look back to the long history of Li-Yue in the East, We can find an hidden historical trajectory, which anticipates the modernization and nationalization which was about to come, long before the wester influences.

2.4 The historical trajectory: Centralisation, Bureaucratization and Spectacularisation

From the late **Western Zhou** to the **Han** dynasty (5th century BC – 1st century AD), there was a deep influence from the Legalist school (centrism instead of feudalism, and bodily discipline instead of family ritual). The **Tang** dynasty (7th–10th century AD) saw internationalization in terms of court style and diplomatic music. The **Song** dynasty (10th–13th century AD) saw the return of Confucian ethics in a series of protocols under nation-wide bureaucratization. **Ming** Dynasty: (14th–17th AD): a state-wide standardized ritual financed mainly by the local government, with series of controls from emperor in practice. **Qing** Dynasty (17th–20th AD): combining Manchu ceremony with Confucian orthodoxy. Great spectacle with minimal interiority. Finally, from **Republic** to **People's Republic**: (20–21st AD): westernization and nationalistic implications, come together with extreme body discipline.

Looking back at history, as long as Yue remained a primary symbol of state legitimacy, each new dynasty carried forward—while also transforming—Li-Yue's framework of its predecessor. From this continuity, one can trace a clear long-term trajectory of **centralization, bureaucratization, and spectacularization** extending over more than 2,500 years. Beginning in the Han dynasty, Li-Yue gradually shifted from a cosmological state rite into a bureaucratic protocol governing bodies, and eventually into a spectacular system manifesting sovereignty. In this sense, the national ceremonies in China today are neither pure historical contingency, nor mere imitations of the west, but the logical outcome of her own history. Any way the history is never deterministic. We will never know what Li-Yue might have looked like without western influences (such as calisthenics, military drills, and march). But it could be safe to say the modern form of Asian state ritual is thus shaped by both western influences and endogenous continuities.

3. LIN-CHI-WEI'S MODULES IN PRACTICE

Yue's Pre-Qin models (ca. 11th–3rd century BCE) operate as educative and generative systems. Through the repetitive practice of Yue, participants regulate personal affect, calibrate their judgment, and allow collective order to emerge in and through embodied practice, rather than merely representing an established order.

By contrast, from the Qin dynasty onward (ca. 3rd century BCE to the 21st century CE), Li-Yue increasingly shifts toward a centralized model. It serves primarily to verify and to display the correct execution of a preconfigured order through bodily obedience and discipline. During this process, the confucian ideal about Yue of individual self-cultivation is reduced to a collective, visualized demonstration of a certain political order defined by it's own elites.

My theoretical inquiry—as well as my “Module Practice” takes shape as a counter-proposition to such established order. It is, in a sense, a reflection on Yue as it has been weaponized throughout the history, and especially through the culture movements in the course of asian modernization.

I am interested in whether, after some three thousand years of Yue practice, and after its full assimilation into nationalist agenda of corporeal training. Might Yue still be able to inspire the potential of generating non-statist protocols? Let's go back to the idea of confucius and make it updated : Li today can still be collectively computed? Embodied social order can still emerge horizontally, from within the community interaction, rather than being imposed from above? And most importantly: Can new rules evolve in real time, shaping constantly by the needs of a changing society, and can itself turn into the motor for the future changes?

3.1 “Inter-human Dynamic Coordination Modules” (IDCM)

These questions have shifted my musical attention away from traditional models of “composition/interpretation” and toward the practice of protocol: not fixed structures, but dynamic rule-sets, iterated and re-generated through live action. Since 2004, I have been developing series of experiments that would later be named “**Human Dynamic Coordination Models**” (HDCM)—systems built from simple rules that combine participants into temporal “human-computer” devices, capable of generating complex and often unexpected sonic phenomena. (see: <http://www.linchiwei.com>)

These so-called “**Sound Models**” include more than Ten basic series, such as:

- **Tape Music** (Mono and Multi-Tracks), inspired by the mechanics of analog tape machines, where audience members act as magnetic heads and speakers.
- **Cellular Automata Music**, based on the theory of Conway's artificial life model.
- **Chapelet Sonore**, where color beads functions as a triggering signal.
- **Quipou Sonore**, A 100 meter long Quipou transfered Incan knot systems into “digital sound notation”.
- **Calligraphie Sonore** , A graphic score with Chinese calligraphy embroidered on a black ribbon, which roughly represent the evolution of pitch, volume and phoneme. The tuning and tonality remained undefined, the singers are obliged to react to one another's sound in order to find their own role in the forming of an unknown harmony.

• **Balloon Music**, A series of instruments attached with a balloon were played by the assistants on stage, then passed to the audience, who were obligated to listen to the others and start the concert for themselves.

If the state's version of Yue reduces individual bodies into screws in a synchronized mega-machine—focusing solely on the correct execution of rules, culminating in a display of order—then I try myself to image alternatively:

A kind of sound protocol in which individual bodies are conditions for each other, interferences for each other, even their reactions can contribute to the next generation of evolving protocols. This idea leads to a sound practice without central conductor, without pre-determined score , where rules and rule-breaking coexist, and where overflow, deviation, and improvisation are not bugs but preconditions of emergence. Here, rules exist only through embodied actions , and the value lies not in the perfection of execution, but in the process of being enacted and tested.

These models have no strictly centralized rhythm, Structures arise not through total control, but through local feedback, lateral calibration, and emergent coherence. Participants form micro-groups keeping on self-correct in real time, and establish local tacit agreements (through sound), the completion of a piece is to certain extent impossible.

In HDCM, protocols are intentionally underdetermined. They do not prescribe detailed modes of expression, such as timbre, duration, pitch, or even precise timing, interaction patterns, or affect—in advance. Individual variation is not a flaw here, and errors are not failures. They are thus treated as signals. Delay, misreading, latency, deviation, and unforeseen (even exaggerated) reactions are structurally permitted. HDCM sets only minimal relational conditions.

The initiators do not “compose” the outcome in the traditional sense, nor do they oversee its progression—there is no conductor at all. Instead, the initiator designs an environment for distributed emergence. HDCM therefore does not claim to abolish power, but to shift artistic organization away from centralized execution toward spontaneous, negotiated, embodied, collective production—where participants generate the work themselves.

The following section offers a brief introduction to two IDCM modules, developed and matured through concrete long-term practices.

3.2 Case Studies in Protocol-Based Sound Modules

Tape Music (2004–)

A long ribbon divides the spectators into a spiral formation. Without any prior instruction, spectators

received a ribbon from initiator and mechanically transfer it to the next begin to transfer the ribbon clockwise, some starts reading or chanting the meaningless syllabic sequences embroidered on it. From these overlapping, not perfectly synchronized vocalizations, a dense acoustic mass gradually emerges—composed of many sub-patterns, it builds itself in volume and complexity, and it is slow, organic, and continuously evolving.

From a cybernetic perspective, traditional Western rehearsal works a parallel signal transmission model: each performer receives the same score simultaneously, and get synchronized through a conductor. **Tape Music**, by contrast, uses a serial transmission model (as described by researcher Tsai Hai-En): the “score” moves from person to person. It functions like a 1970s tape delay effector—or perhaps more accurately, like a live mechanical canon.

Each participant hears their neighbors first, entering the system based on the rhythmic patterns they perceive, which differs in different position. This moving framework gives individuals significant autonomy, regarding to pitch, duration and timbre, yet the loosely synchronized framework allows the sub-groups to produce emergent intensifications—patterns that can be perceived by the rest of the participants, yet remain unstable and keeping evolving. Much like how bees follow collective pheromone trails to form hives, participants collectively reinforce certain sonic patterns, producing an effective form of collective computation.

Notably, Tape Music is not a singular work. It has generated over twenty versions in the past 20 years—each evolving like software, with beta releases, updates, and full revisions. Other artists and social groups have adopted the structure in different contexts: and used them in wedding, party, therapy workshops, youth camp etc, and it became a staple in UK choir **Musarc**'s repertoire since 2016.

	Classic Music Rehearsals	Interhuman Dynamic Coordination Modules (Example: Tape Music)
Conducting	Conductor-Centered Tempo, volume, timbre, and balance are centrally controlled by a single conductor.	Decentralized Structure Volume, timbre, and melody are self-determined by each participant, who independently balances their sound or forms collective patterns with others
Synchronization	Following the Conductor's tempo – Strict Synchronization	The group synchronizes with ribbon symbols rather than a conductor's gestures, allowing for flexible timing.
Score	Individual Scores for Everyone	Equal Roles or Self-Defined Division of Labor
Labor Division	Divided Roles Assigned by the system	All Expressions Welcomed – Embracing diverse and distinct sound expressions.
Expression	Suppressing Individual Sound Characteristics—Divergent personal expressions are not encouraged.	Imaginative Instruments – reimagining power dynamics within existing ones.
Social	Fulfilling Established Roles Within Society	Imaginative Instruments – reimagining power dynamics within existing ones.

Table 1. A comparison to traditional music structure

3.3 Cellular Automata Music (2006–)

Thirty-six participants enter sequentially, forming a 6×6 grid. Each of them act as “cell” within a live system governed by the rules of cellular automata. A signal—such as a wooden fish strike or a flashing light—marks an Act. Before the first Act, each participant selects a small instrument from four provided instruments.

From the second Act onward, each participant's (next) action and choice of instrument is determined by the status of their four immediate neighbors (front, back, left, right). The rules are as follows:

- If your instrument differs from all your neighbors, you can continue to play it during three more Acts, then stop permanently (died).
- If it matches only one or two neighbor, switch instruments in the next Act.
- If it matches three, continue playing the same instrument in the next act—but if the same instrument is played for more than three consecutive Acts, stop permanently (died).
- If all neighbors share your instrument, stop immediately in the next Act.(died)

Participants need to listen to what the other “cells” are doing thus determined what to play, simply because there is no musical instruction but a sequence of instrument choices with slow diminution (of instruments). This produces a gradually evolving sonic system, capable of shifting sound textures. The same principles can be adapted to choreographic structures for group movement.

4. CONCLUSION

This paper has argued that pre-Qin Li-Yue is best understood not as a work tradition but as a **protocol tradition**. Rather than organizing sound toward the realization of autonomous musical works, Yue employed embodied procedures through which sound coordinated bodies, affects, and social relations. Its primary concern was therefore not musical representation but the operational generation of normative order.

Viewed from this perspective, HDCM is neither a reconstruction of Li-Yue nor an attempt to revive its historical moral or political agenda. Rather, Li-Yue serves as a theoretical coordinate from which contemporary sound practice may be reconsidered. If modern music has largely inherited a work-centered ontology, HDCM investigates the possibility of a **protocol-centered one**.

Here, composition shifts from prescribing sonic outcomes to designing the minimal operational conditions under

which collective organization may emerge. Protocols remain intentionally underdetermined. Rather than specifying sounds, they specify relations. Participants negotiate timing, response to each other, interact to form local decisions - sometimes even unconsciously, while delay, mis-interpretation and deviation become productive elements of the system rather than failures to be corrected. Musical structure is therefore not realized in advance but continuously enacted through collective computation.

The question is consequently no longer how to compose an ideal musical work, but how to design **operational protocols** through which new forms of collective organization may emerge. HDCM does not propose a new musical style, nor an updated version of Li-Yue. It proposes **protocol as a new compositional object**: one in which artistic practice shifts from producing works to designing the conditions under which societies may experiment with their own modes of coordination.

Acknowledgments

The author would like to thank Ang Lu-Teng, Tsai Hai-En and Wolfgang Schwabe for their valuable feedback. Proofreading by Wolfgang Schwabe.

5. REFERENCES

[1] *Notes on Li-Yue in Confucius School* (聖門禮樂誌) Qilu Publishing House, by Kong Lingyi pp. 101-152, 2019

[2] Radio Gymnastic Exercises as a Device and the Colonial Taiwan -Journal of Cultural Studies, Pei-Hsien Hsu, pp161-199, 2011

MUTABLE SCORES: EMBRACING NON-FINALITY IN NOTATION

Alexandre Sasset-Blouin

Université de Montréal

CIRMMT student member

alexandre.sasset-blouin@umontreal.ca

ABSTRACT

In Western art music, a work is commonly understood as a discrete and final entity, instantiated through a score or fixed instructions that govern its recreation. Although various practices experiment with a score's determinacy, few musicians have envisioned a work's notation as a continuously evolving and perpetual process. Drawing primarily on an analysis of Lydia Goehr's work-concept, this paper first examines how works are largely perceived as finished entities. It then presents Lauren Redhead's iterative compositional approach as a case study of scores deliberately designed to resist finality. Building on this framework, the paper introduces the mutable score through the examination of a composition currently in development. The proposed model supports perpetual adaptation, reworking, and recontextualization, treating iterations not as derivatives but as active components of an ongoing compositional process and traceable records of the work's evolution. The paper further considers how such iterations may be documented, preserved, and circulated to remain operative in subsequent realizations.

1. INTRODUCTION

In *The Imaginary Museum of Musical Works*, philosopher and musicologist Lydia Goehr articulates the work-concept: the conditions and ideals that regulate Western musical practice and the creation of musical works [1]. In her broader examination of the work-concept, Goehr offers the following account of the perception of musical works in the centuries prior to its emergence around 1800: "Music was not always produced to outlast its performance or survive more than a few performances. And when it did survive many performances, numerous changes could and usually would be made to the music in the process. Rarely did musicians think of their music as surviving past their lifetime in the form of completed and fixed works. When musicians thought about repeatability, they thought more of the multiple uses of themes and parts for various different occasions, than of one and the very same whole composition being repeated in performances dedicated to the performing of that very composition." [1, p.186] According

to Goehr, the emergence of the work-concept established a clear separation between the musical work and its performances, framing the work as a discrete and final entity.

Many experimental practices and ontological discussions engage with the level of determinacy of a work and the variation among its individual representations. However, even in cases of maximal indeterminacy, the work's notation is still largely understood as a definitive set of instructions for recreating a finalised composition.

Drawing on a creative process developed by Lauren Redhead, this paper seeks to further engage with finality by imagining the score¹ as an ever-evolving and perpetually unfinished element of a work. From there, it proposes the mutable score as a new approach to composition and notation, where non-finality and iterative transformation are integral to the work from its inception.

2. THE FINALITY OF WORKS

This section argues that finality is ubiquitous in the Western perception of musical works, and that the work-concept's pervasive influence extends even to practices that exist outside its framework or explicitly seek to challenge it.

2.1 Ontology and authenticity

Firstly, to consider a musical work as final, it is necessary to distinguish the work itself from its various instantiations. The finality of a work does not imply uniformity across performances, recordings, or other realizations. In *The ontology of musical works and the authenticity of their performances*, Stephen Davies [2] contends that multiple performances of a musical work, even if they vary greatly in how they are realised, can still be considered realizations of a singular work. In fact, as Davies points out, an authentic realisation of a work may involve significant variation between instances², and even inauthentic realisations which vary by misrepresenting the work in some way, are still perceived as the same work [2]. Davies also includes adaptations or score modifications, such as re-scoring a piece for different instruments, by noting that such performances "may be recognisable as a performance of a particular work even if it is not possible to recover all of the work from the performance" [2].

Copyright: ©2026 Alexandre Sasset-Blouin. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ In this context, the term score is used broadly. It refers not only to traditional notation but to any set of material instructions that indicate how a composition may be realized.

² Davies gives the example of a baroque piece with figured bass. The figured bass may be executed in diverse ways across instantiations; nonetheless, other more determined features maintain their identity as a single work.

Davies' analysis emphasizes that the musical work exists independently of any particular instantiation, a perspective reflected in Goehr's articulation of the work-concept. This independence allows the work to retain its status as final, even in the absence of a definitive or "true" version of its performance. Thus, a musical work and its score is generally considered finished and complete, despite variations in interpretation.

2.2 The pervasiveness of the work-concept

The final chapter of *The Imaginary Museum of Musical Works* describes the pervasiveness of the work-concept: "the romantic aesthetic originally provided has continued, since 1800, to be the dominant view. This view is so entrenched in contemporary thought that its constitutive concepts are taken for granted. We have before us in fact a clear case of conceptual imperialism" [1, p.245]. As Goehr notes, musical practice is interpreted through the framework of the work-concept, even when the composer did not conceive it in those terms. Through "conceptual imperialism", the ideals of the work-concept then infiltrate practices that do not traditionally operate within its framework.

Jazz music, for example, does not typically adhere to the work-concept, foregoing the score and the work in favor of the chart and the standard. Yet, according to Goehr [1, p.250], some jazz musicians adopt the "institutional conventions" of the work-concept, to achieve the perceived authenticity of "serious" music. Michael Talbot similarly identifies a broader shift within Western musical culture from performance-centered practices toward composer-centered ones [3]. Both authors point to the professionalization of the composer and the commercialization of music as forces encouraging the conception of works as distinct, authored, and complete entities capable of repeated realization by performers [1, 3].

Even the most experimental approaches that seek to explicitly question and disrupt the norms of this model are paradoxically engaged with it. John Cage's work *4'33"* (1952), mentioned by Davies, Goehr, and Talbot, provides a compelling case study that appears to deliberately depart from the ideals of the work-concept. While the work seeks to relinquish all compositional control, challenging "formal notions of pre-composition, fixed instrumentation, repeatability, and compliant performance" [1, p.262], it fails to completely detach from the mainstream perception of a work as finished entity. For a range of reasons, whether professional and commercial pressures tied to the recognition of the composer or the force of institutional conventions, in *4'33"* and other similar works "the formal constraints of the work-concept have ironically been maintained" [1, p.263]. Tellingly, the score was published by Edition Peters in a conventional, typewritten format, first in 1960 and again in 1986 [4].

As illustrated above, the ubiquity of the work-concept allows experimental works such as *4'33"*, which almost entirely remove composer intervention, to be understood as final, authored, and repeatable. As will be discussed in the following sections, even where works engage with deter-

minacy in unconventional ways, the condition of finality is largely maintained.

2.3 Determinacy and finality

Contemporary works and notations operate across varying degrees of determinacy, some allowing for maximal liberty in the performer's interpretation while not straying into the purely improvisational. Goehr remarks that the composer's intent, and the performer's faithfulness to it (*Werktreue*) are the differentiating factors between the performance of a work and an improvisation: "There is a significant difference between performers who comply with a notation which specifies that they play what they will for as long as they want, and in so doing perform the composer's work, and performers who improvise freely around themes because that is what it means to produce their sort of music. What is absent in the second case, but what is present in the first, is the performer's commitment to *Werktreue*." [1, p.263] This illustrates the flexible nature of the work-concept regarding the determinacy of realizations, allowing for even the most open and free forms of notation to be considered finished and authored entities, of which each performance or recording is an instantiation.

On one extreme of the determinacy spectrum, the performances of open scores, such as Cornelius Cardew's classic work *Treatise* (1967), could be wholly unrecognizable from one instance to another, yet would be considered as instances of a singular finished work represented in a fixed notation. In the introduction to *the Treatise Handbook*, Cardew explicitly distinguishes between the completed score and the accumulated experiential knowledge of the work: "However, the years of work on *Treatise* have furnished me with a fund of experience obviously distinct from the experience embodied in the score itself. And this fund continues to accumulate, since my experience of and with the piece is by no means completed with the completion of the score;" [5]. While *Treatise* and similar scores bring into question what proportion of the work resides in the score or in each individual performance, the work is still finalized through its completed notation. This example demonstrates that the work-concept's condition of finality can encompass even maximal variation between instances, indicating that finality cannot be challenged solely through indeterminacy.

3. NOTATION AS A PERPETUALLY ITERATIVE PROCESS: THE CASE OF LAUREN REDHEAD

As demonstrated by Davies and Goehr, the condition of finality is evident across a broad range of works, regardless of determinacy or attempts by composers to resist the conventions of the work-concept. While certain practices attempt to remove as much composer control as possible, exploring the limits of determinacy, finality, defined here as the understanding of a work and its score as a finished entity, of which each performance or recording is an instantiation, is more rarely probed. As Talbot and Goehr suggest, professionalization and commercialization certainly play a role in the need for even the most avant-garde composers

to adhere to the work-concept, and to be seen as serious composers of completed musical works [1, 3]. One contemporary composer who deliberately addresses finality in her work is Lauren Redhead, whose compositions will be examined as a case study to inform the subsequent discussion of mutable scores.

Redhead engages directly with the notion of finality through her approach to the composition and notation of two of her works: *ijereja* and *entoptic landscapes*. In these works, composition, notation, and interpretation exist as iterative processes that could extend indefinitely. The scores of these works are continually re-notated and directly include elements from previous iterations: “In the case of each piece, materials created in and by previous performances go on to inform future performances of the music. As such, there can be no ‘definitive’ performance or statement of the works, and nor can they ever be considered finished or bounded.” [6]

According to Redhead, notation in the traditional work-concept is understood as a set of instructions that allow for the recreation of a work, and a material preservation of the musical idea³: “As such, notation is always outside of the work, as a form of documentation, rather than inside the work, as a part of its creation or re-creation.” [7] Her approach to composition, which she describes as a perpetual “re-performance” of notation, repositions the score as a “work (labour) of notation and not as a representation” [7]. By including notation in the ongoing work, an active and living element that is recreated iteratively, Redhead’s works engage with the finality condition of the work-concept. The notation is not only adapted or modified to fit different contexts or instruments, but part of a perpetually unfinished work.

In the same way that a score’s interpretation evolves over time—each performance and recording imparting nuances that shape the musician’s understanding of how the piece might be played—the scores for *ijereja* and *entoptic landscape* are marked by traces of their past iterations. These traces include not only fragments of past versions of the score, but also elements adapted from recordings of performances integrated into subsequent iterations of the notation. The nonstatic score and its evolving interpretation become mutually dynamic, each continually impacting the realization of the other.

While describing *entoptic landscapes*, Redhead reflects on the work’s renewed status under the work-concept: “In this way, the piece can both be considered complete and expanding: it reacts to its performers, performance circumstances, and internal contradictions in an ongoing series of iterations.” [6] Despite their lack of finality, *entoptic landscapes* and *ijereja* are works. Just as performances or recordings are considered instantiations of a singular work, the scores can be similarly understood. Interestingly, this approach allows the same work to exist in different forms, including determinate or indeterminate notation, or even installations or other formats⁴.

³ Redhead cites Christopher Williams’ “prescription-preservation model of the score”

⁴ Redhead, on *ijereja*’s multiple formats: “Today, the piece encompasses several live acousmatic and multi-channel performances, a studio

Finally, the works are cumulative. As Redhead explains while discussing *ijereja*: “This process is one of accretion, in that more materials of all types may be subsequently drawn into the project without changing its identity.” [7] While all works intended for multiple recreations accumulate performances or recordings over time, these two examples extend this accumulation to include the scores themselves. These materials are not separate streams of iterations but influence one another in an evolving network. In her essays on the works, Redhead does not delve into how or if the material trace of each iteration is preserved, or if they are kept distinct from one another. While the works include all past materials accumulated to that point, it is not clear if she considers each iteration as a distinct snapshot of the work, or if the work is more of a palimpsest, each new addition overlapping the previous and obscuring the separation between layers.

My proposed approach, developed throughout the rest of this paper, similarly envisions a work that is perennially non-final. It adopts Redhead’s accumulation process but sets itself apart by emphasizing and preserving individual iterations as active elements in a perpetually evolving work.

4. TOWARDS A MUTABLE SCORE

I propose the mutable score as my own approach to notation and composition, designed to adapt and evolve indefinitely, each generation building on the last, inheriting some traits while abandoning others. In parallel with Redhead’s approach, it engages with notation within the work-concept by suspending the condition of finality typically associated with it, while highlighting the ongoing relevance of each iteration. This section opens with a description of a score currently in development, created to test the mutable score approach, and then uses this experimental score to illustrate three core components of the method.

4.1 Mutable experiment

mutable experiment is an ongoing work created to explore a perpetually evolving form of notation. It functions as a tool to test and develop the mutable score approach through practice-based research. The objective is to enable the notation to evolve over multiple iterations, adapting to different contexts and intentions, while preserving a material trace of each stage. In this way, each produced score continues to exist as a living element of a perpetually non-final work that can be reused in further versions or consulted to inform further compositional decisions.

The first iteration of the score, of which an excerpt is pictured in Fig. 1, was created with a Max for Live device currently in development at the University of Montreal’s Laboratoire formes • ondes⁵. The device allows users to produce scrolling, animated graphic scores directly in Ableton Live by assembling PNG images of symbols and shapes. The scores are then played back to the musicians

album, multi-modal gallery presentation, audio-visual installation, scores for specific instruments and open notation.” [6]

⁵ More information can be found here: <https://lfo-lab.ca/notation-performative-en>

to guide performance. The composition of the first iteration was guided mainly by visual aesthetics, using an open form of notation to allow a large amount of flexibility in interpretation. Since the score was expressly created to evolve, this approach was selected to encourage a multitude of musical ideas to emerge, which could then inform the next generations. The graphic form was chosen to allow a multitude of instruments to be used, including my own modular synthesis and feedback-based systems.

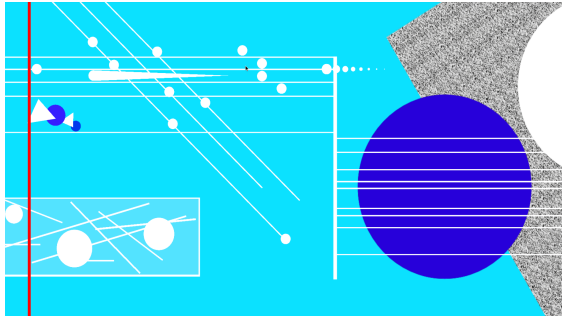


Figure 1. excerpt from the first iteration of *mutable experiment*.

The first version of the score was performed in a studio session, with an experienced classical guitarist on electric guitar and pedalboard, alongside my performance on a no-input mixer system and modular synthesizer. Little instruction was given prior to the session, and multiple different interpretations were recorded. Constraints included a fixed length of 3 minutes, and a scrolling playhead (pictured in Fig. 1 and Fig. 2) to allow the musicians to follow the score at the same pace for each repetition. This proved useful for the subsequent examination of how the symbols were interpreted, allowing a direct comparison of different interpretations.

The recordings from this session were not intended for publication, but rather to inform the next generation of the score. An arrangement was created from different segments of these recordings, from which another graphic score (see excerpt in Fig. 2) was developed using the same score creation and playback device. This new score inherited certain traits from the previous version while discarding others, yielding a more determinate score intended to more closely reproduce the assembled materials.

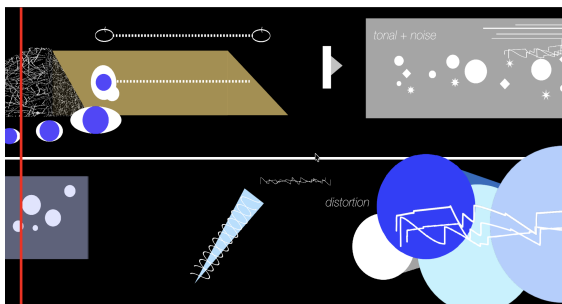


Figure 2. excerpt from the second iteration of *mutable experiment*.

This score was also performed in a studio session to ob-

tain recordings that will inform future generations. From this point, the score will diverge into two distinct versions: a graphic score in open notation allowing free interpretation of symbols, and a more determinate score combining traditional notation with graphic scoring methods (see Fig. 3). This generation is currently in progress.

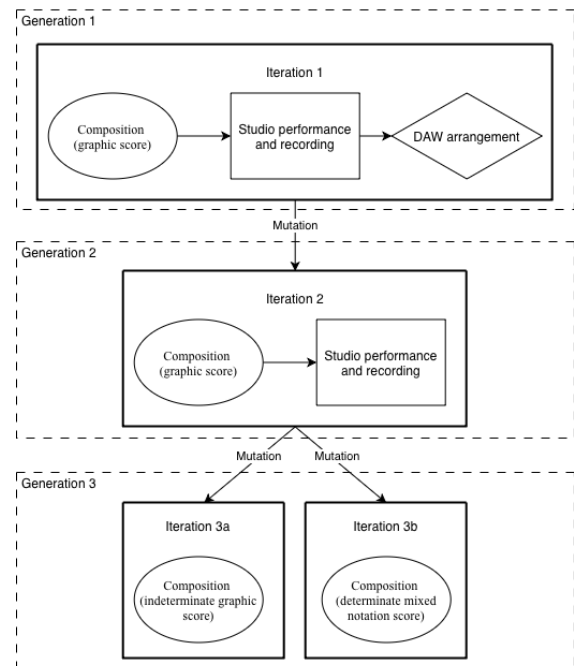


Figure 3. Diagram of the first three generations of *mutable experiment*.

It is important to note that, in keeping with the mutable score approach described in the following section, each version of the score is preserved as a living component of the work as a whole. Accordingly, each iteration of the score should not be understood merely as a step toward a finalized version, but as an ongoing element that may be re-performed in its own right, transformed into divergent versions, or consulted as a trace of the work's evolution.

4.2 Characteristics of the mutable score

Drawing on *mutable experiment*, this section outlines three principal characteristics of the envisioned mutable score approach.

4.2.1 Mutation

Mutation in the biological sense is one of the primary methods by which species evolve, allowing for the emergence of new characteristics that can be conserved through natural selection. The mutable score evolves in a similar manner, retaining or discarding traits through generations in response to the composer's intentions and needs.

This process reflects R. Keith Sawyer's conception of creativity as iterative and improvisational, involving continual movement between its various stages, rather than a linear progression from idea to execution [8]. A composer of a new musical work does not immediately produce a fully

formed and definitive idea, which is then passed on to musicians who execute the instructions to the best of their ability. Instead, the composer engages in a recursive negotiation between the musical idea and its realization. Henrik Frisk and Stefan Österjö [9], describing the process of adapting a musical idea for the harp, offer an account of this process in action. First, the initial musical idea is transcribed into notation but is unplayable on the instrument. Working with a harpist, multiple iterations of the same musical idea are tested. Changes (mutations) are retained or discarded between iterations until a balance is reached between playability and adherence to the original idea.

In contrast to the work discussed above, where iteration and selection shape a musical idea and score into a final form, the mutable score is allowed to evolve indefinitely, incorporating new ideas as spontaneous creative developments or reactions to the needs of new presentation contexts and formats. This process would differ from a simple adaptation, where the musical idea is modified to allow it to be played as faithfully as possible on a different instrument, such as a piano reduction of an orchestral work. Instead, the mutable score is conceived to not only adapt to the parameters of a new presentation, but to permit the modification of the musical idea itself as well as its instructions for recreation. The *genetic material* of the work — the music, the score, the form — can thus progressively evolve, yielding instances with high degrees of variation, yet still tied to the same work by a common traceable genealogy.

In the case of *mutable experiment*, mutation manifests in multiple forms. Studio recordings enabled the identification of score elements whose interpretations were deemed effective and subsequently carried forward into the next iteration. Fig. 4 illustrates the evolution of different symbols from generation 1 (blue and white) to generation 2 (black, grey, and white).

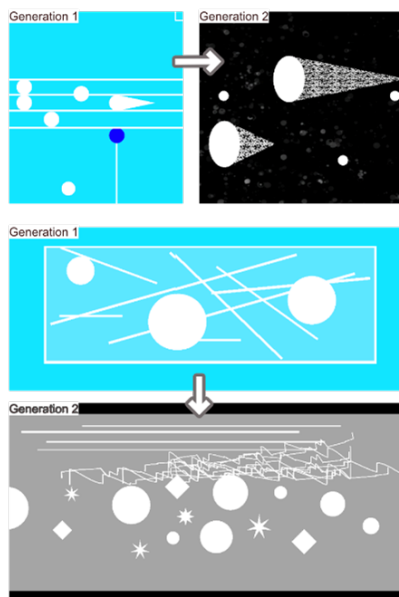


Figure 4. evolution of symbols from generation 1 to generation 2 of *mutable experiment*,

In the third generation, mutation is expressed through the evolution of the score from an indeterminate graphic notation to a more determinate mixed notation. As illustrated in Fig. 5, the continuity between symbols across generations remains evident, even as the score evolves not only in content but also in form.

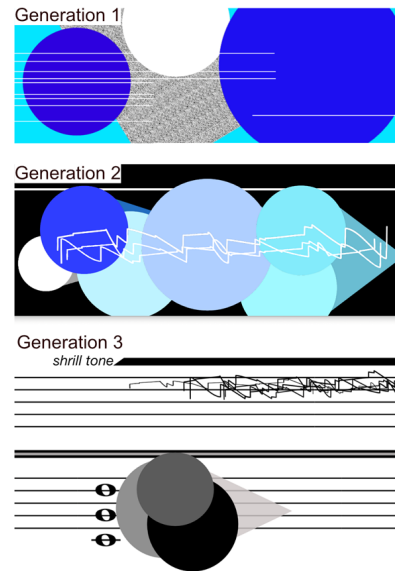


Figure 5. evolution of a symbol through three generations of *mutable experiment*,

4.2.2 Phylogeny and genetics

The perpetual mutation of the score allows, as Redhead describes, the “re-performance of notational labour”, inserting notation “inside the work” [7]. However, the mutable score differs from Redhead’s approach by its emphasis on the genealogy of the score and the record of its evolution. Phylogeny, the study and representation of evolutionary history, provides a useful analog for the hereditary and iterative nature of the mutable score. Through a phylogenetic approach, the evolution of the score can be traced and analysed throughout its iterations, allowing the work to exist as a single continuous entity with multiple evolutions and divergent points. Even if progressive variation produces seemingly unrelated instantiations, their lineage can be traced to common points of divergence.

Taking *mutable experiment* as an example once again, although early generations maintain significant similarities, successive evolutions could produce scores that share little with the earliest versions. The diagram of its iterations (Fig. 3) would extend through evolutions and divergences, resembling a phylogenetic tree, as depicted in Fig. 6. In this case, the identity of the work as a singular continuous entity is assured by its genealogy and the traceable link between each generation. The work can thus be understood as a continuous compositional process rather than a series of drafts leading to a final proposition.

Nicholas Donin’s work, which includes genetic analysis in music studies, highlights the potential of preserving a material trace of a work through its compositional

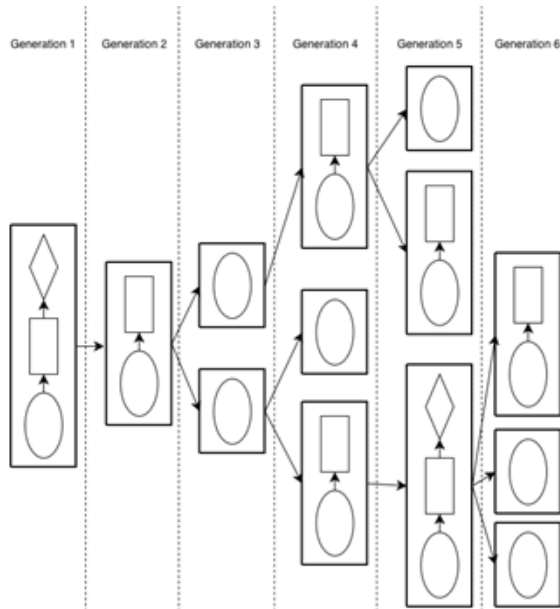


Figure 6. Phylogenetic tree of *mutable experiment* including potential future generations,

stages. According to Donin, musicology alone seems insufficient to establish composition as an autonomous object of study [10]. Including genetic analysis—the study of a composer’s creative process and its material artifacts, such as notes, drafts, and score versions—offers new perspectives on works as well as new creative possibilities. Through this approach, Donin emphasizes the compositional process as an integral part of the work, a perspective that aligns with Redhead’s iterative composition and the mutable score approach outlined here. In discussing a research project on the genetic analysis of Philippe Leroux’s 2002 composition *Voi(rex)*, Donin explains how the work’s genetic record was used to inform the creation of multimedia devices that highlight listening traits noted by the composer and also led to a new composition by Leroux reusing ideas, materials, and techniques from *Voi(rex)* [10]. Notably, this evolution of the work into two divergent iterations echoes the ambitions of the mutable score.

4.2.3 Preservation

As mentioned previously, the material traces of each iteration of the mutable score must be preserved to allow the tracing of its evolution through documentation and chronological organization. Documentation need not capture every incremental change; instead, it should emphasize key stages or critical divergences.

Preserving the different iterations is a crucial aspect of the mutable score. Whereas the evolution of a traditional musical work must often be painstakingly reconstructed from the analysis of a composer’s notes and comments⁶, those involved in creating a mutable score are responsible for maintaining its genetic record from the very outset of the process. Preservation is what allows the mutable score to function and what differentiates it from other forms of

composition and notation, allowing previous iterations to remain useful over time and maintaining the status of the work as a continuous entity. Thus, preservation, like notation, is incorporated into the ongoing performance and labor of the work, as an active part of the compositional process, with composers shaping the genetic record through selection and documentation.

Multiple strategies could be employed for preservation, facilitated by modern technological advances in documentation and storage. While they will not be discussed in detail here, the field of software engineering could be rife with potential solutions, as preservation of data through iterations is critical to modern software development practices⁷.

5. COLLABORATIVE MUTABLE SCORES

The mutable score concept is still in its early stages and, for now, is confined to my own practice while it is further developed. In the long term, it is envisioned as a means of fostering and facilitating collaboration, becoming an open-source project available to a broader community. In the intended implementation, the entirety of the work, including all iterations, will be made publicly available, allowing anyone to continue to iterate on and evolve the work as they see fit. By encouraging contributors to engage with an evolving work and freely adapting the score and its materials, a single mutable score could diverge into autonomous branches, with the expectation that such modifications be documented as part of the work’s ongoing development. Here again, parallels can be drawn with software development, where open-source models encourage collaboration and reappropriation, while preserving authorship and agency over one’s work⁸.

6. CONCLUSION AND NEXT STEPS

This paper outlines the conceptual and methodological scaffolding on which mutable scores can be built. It demonstrates the ubiquity of finality within Western music practice to propose a new method of scoring and composition that explicitly seeks to resist it. This model, in addition to challenging finality, highlights the potential afforded by the documentation and preservation of the material trace of a work through its evolutionary stages. Ultimately, it proposes to expand the conception of a work to include notation and preservation as integral and ongoing processes, continuously reperformed as part of the labor of creation.

As mentioned previously, this research is in its early stages and is currently in development through a practice-based research approach. *mutable experiment*, as well as new mutable scores, will undergo continuous iteration, being adapted to different instruments and forms. Eventually, all scores and other materials will be published as an open-source project and made available to composers or musicians who wish to interpret, adapt, or transform the work.

⁷ See Software Evolution by Kim et al. [11].

⁸ Multiple types of software licensing strategies (such as copyleft practices) could help composers maintain their desired level of control over their work.

⁶ Donin gives the example of Beethoven’s notes in the margins of his scores or the collections of his sketches.

A preliminary step will be to adopt a system to organize, preserve, and make the material accessible. As discussed above, multiple solutions intended for software development could be used. An iteration hierarchization system must be developed to determine when a modification constitutes an adaptation and when it represents a divergence substantial enough to warrant a separate branch. Likewise, the specific materials required for each individual iteration have yet to be determined. Furthermore, questions of authorship and ownership will have to be addressed, although open-source software practices, as mentioned, may offer solutions. These questions will become clearer over time, particularly if community engagement allows the project's methodology to evolve in tandem with the works themselves.

Finally, an artistic presentation of a mutable score should be conceived to communicate the principles and ambitions of the approach. Rather than performing only the most recent iteration, the exposition should make the score's evolution and multiple versions clearly apprehensible, allowing the audience to experience the process of adaptation and divergence. This could take the form of consecutive performances of different iterations and include an alternative format, such as a sound installation, that emphasizes the work's iterative and evolving nature.

7. REFERENCES

- [1] L. Goehr, *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Oxford: Clarendon, 1992.
- [2] S. Davies, "The Ontology of Musical Works and the Authenticity of their Performances," *Noûs*, vol. 25, no. 1, pp. 21–41, 1991.
- [3] M. Talbot, "The Work-Concept and Composer-Centredness," in *The Musical Work: Reality or Invention?*, M. Talbot, Ed. Liverpool University Press, May 2000, pp. 168–186.
- [4] J. Cage, *4'33"*. Edition Peters, 1960.
- [5] C. Cardew, *Treatise Handbook*. London: Edition Peters, 1971.
- [6] L. Redhead, "Entoptic landscape and ijereja: Music as an iterative process," *New Sound International Journal of Music*, vol. 49, pp. 97–113, 2017.
- [7] L. Redhead, "Iterative Processes: Notating, Composing, and Performing as Journey Forms," *Contemporary Music Review*, vol. 41, no. 2-3, pp. 266–280, 2022.
- [8] R. K. Sawyer, "The iterative and improvisational nature of the creative process," *Journal of Creativity*, vol. 31, p. 100002, 2021.
- [9] H. Frisk and S. Östersjö, "Negotiating the musical work:," in *International Computer Music Conference, ICMC 2006*, New Orleans, United States, 2006, pp. 242–249.
- [10] N. Donin, "Quand l'étude génétique est contemporaine du processus de création : nouveaux objets, nouveaux problèmes," *Genesis*, no. 31, pp. 13–36, Oct. 2010.
- [11] M. Kim, N. Meng, and T. Zhang, "Software Evolution," in *Handbook of Software Engineering*, S. Cha, R. N. Taylor, and K. Kang, Eds. Cham: Springer International Publishing, 2019, pp. 223–284.

(RE-)DEFINING THE POETIC TEXT SCORE

Craig Pedersen

Edith Cowan University
c.pedersen@ecu.edu.au

ABSTRACT

Text scores emerged within post-1960s Experimental music as the musical score's definition expanded beyond the representation of foreseeable or reproducible musical outcomes. Using language in place of other notations, text scores employed indeterminacy and openness to establish an acceptable range of possible realisations. While instructional text scores—those that relatively clearly indicate musical action—have been widely discussed, their more ambiguous, non-instructional counterparts have received comparatively little attention. This paper addresses this gap by proposing a redefinition of such works as *poetic text scores*, framing them not by the absence of instruction but through their ambiguity and poetic function.

Drawing on examples from Fluxus-associated artists, Pauline Oliveros, the Wandelweiser collective, and contemporary experimental practice, the paper surveys a range of poetic text scores, categorising them according to word use, spatial layout, and thematic qualities. Building on post-Saussurean structuralist semiotics, the paper advances a model of the score as an object signifying acts in time, with the act of interpretation arising to answer the questions “which acts, and when?” This framework is applied first to instructional text scores, before turning to Roman Jakobson's theory of the poetic function to clarify and account for the poetic text score's identity as score.

1. KEY QUALITIES OF POETIC TEXT SCORES

Within music, a text score can be defined as a score using the written word to convey compositional ideas in place of other forms of notation [1]. They have been described as a place where language can subtly and effectively be applied to sound [2], using qualities of both language and music to artistic ends [3]. Although the oral or written transmission of musical ideas is present within many genres, such as rock, jazz, or folk musics, this paper focuses on scores written in alignment to the

Experimental¹ music movement borne out of Western classical music. Further, while the use of text in-score as rehearsal or performance markers is not new, we see the medium of text-as-score emerging from the 1960s onwards, largely post-John Cage, with Fluxus-associated artists such as George Brecht and Yoko Ono, composers Pauline Oliveros and Karlheinz Stockhausen, later from composers associated with the Wandelweiser collective such as Antoine Beuger, Stefan Thut, and Manfred Werder, and continued contemporaneously by composers such as Lo Wie, Ryoko Akama, Sarah Hughes, Sage J Harlow and Elizabeth Millar.

Although a variety of terminology and descriptors have been suggested (Table 1), broadly speaking text scores fall into two categories [4], [5], [6], [7]: instructional scores which focus on the use of text to (relatively) clearly communicate a range of actionable musical ideas, and non-instructional scores which are definitively ambiguous in the ways they point to actionable sounding interpretation, the latter of which is of most interest to this paper. Additionally, some scores combine qualities of both instructional and non-instructional scores to form a composite [3], [4], while others use text in addition to other forms of notation.

While to date historical background and defining features of Experimental music and open works have been covered in detail [8], [9], [10] and groundwork has been set in understanding varying forms of notation [7], [11], [12], including a comprehensive study of the specific ways many, primarily instructional, text scores use language to achieve their aesthetic ends [1], missing from the discourse is comprehensive discussion of the nature of non-instructional text scores.

1.1 Terminology

Non-instructional text scores are a form of open work, a term which conventionally describes compositions for which multiple sounding outcomes across different performances is an aesthetic feature, achieved through a variety of compositional techniques. As such, using ambiguous text in place of other notations, non-instructional text scores are likewise ambiguous in the ways they point to actionable sounding interpretation, aestheticising and featuring multiple interpretations and

Copyright: © 2026 Craig Pedersen. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ Capitalised here and throughout to distinguish between the Experimental movement borne out of Western classical music, and experimental music in general.

sounding difference between performances. Accordingly, they have been described as “suggestive” [6], “mysterious or enigmatic” [5], or “poetic” [6], [7]. Anderson [4] suggests the terms “allusive, (conceptual or ambiguous)” for text scores that do not contain clear instructions referring to action, using as example La Monte Young’s *Piano Piece for David Tudor #3* (1960) [13] reading “Most of them / were very old grasshoppers”. Lely & Saunders [1] discuss grammatically “moodless” scores that place words in such a way that they lack command, statement or question, thus creating a marked lack of clarity, drawing on scores from George Brecht’s landmark work *Water Yam* (1963) [14] collection as example; Montgomery suggests [3] “poetic” and “para-poetic score” to denote those using qualities of both language and music, focusing on the ways in which the text-based works of Michael Pisaro and Manfred Werder may or may not make clear how to be translated into sound.

Instructional:	Non-instructional:	Composite:
Action score	Ambiguous	Hybrid
Event score	Allusive	Composite
Instructional text score	Conceptual	Para-poetic
Prose score	Mysterious	
Prose notation	Non-instructional	
Text-based action score	Poetic	
Text notation	Suggestive	
Verbal notation		
Verbal score		

Table 1: terms used to identify and describe text scores

My preference is for the term *poetic* text scores, as the term positively describes qualities and characteristics of said scores, as opposed to the term non-instructional, which defines them by qualities absent. Discussing the nature of poetry and ambiguity, sociolinguist Roman Jakobson [15] states that “ambiguity is an intrinsic, [...] corollary feature of poetry” (p. 370). As such it has been argued that ambiguity is pragmatic, in that intended meaning left unclear to the reader [16]. Following this logic, although a detailed account of their poetic nature follows, as a baseline we can take as given that what defines poetic text scores is not simply the use of suggestive, mysterious, allusive, or ambiguous language, but that their ambiguity extends to their musical meaning, definition, and interpretability as scores.

2. CATEGORISATION OF POETIC TEXT SCORES

Given the use of ambiguous language as relating to their actionability as scores, poetic text scores are difficult to categorise in the same way as other open scores, unable to be clearly categorised according to their formal structures, application of parameters or additional forms of instruction. Further, given their variety they cannot clearly be categorised by type of language used and there are not readily present examples of scores containing

characteristics of different genres of poetry, such as sonnet, epic, or narrative. In most cases, the closest correlation to poetic genre is that of free verse and concrete poetry, and occasionally haiku, a finding congruent with the understanding that the poetic nature of such text scores depends primarily on the ambiguity of their actionability as pertains to their nature as scores. As such, a simple approach to categorisation is to group them according to their word use and placement²

Single words or ideas	• <i>October 24, 2022</i> (2022) [17] by Craig Pedersen • <i>Hide Piece</i> (1961) [18] and <i>Cut Piece</i> (1964) [19] by Yoko Ono • <i>Suitcase</i> (1963) [20] by George Brecht • <i>wind</i> (2017) [21] and <i>grass</i> (2017) [22] by Lo Wie
Word groupings or lists	• <i>2005¹</i> (2005), <i>2008³</i> (2008), and <i>2012³</i> (2012) [23], [24], [25] by Manfred Werder • <i>for Strings</i> (1969) [26] by Howard Skempton • <i>Three Aqueous Events</i> (1961) [27] by George Brecht
Short statements	• <i>2010¹</i> (2010) [28] by Manfred Werder • <i>Chair Event</i> (1963) [29] by George Brecht • <i>Earth piece</i> (1963) [30] by Yoko Ono • <i>Piano Piece for David Tudor #3</i> (1960) [13] by La Monte Young • <i>April 14, 2022</i> (2022) [31] by the Craig Pedersen • <i>Sifted Ashes of Air</i> (2022) [32] by Sarah Hughes
Longer verses	• <i>The Earthworm Also Sings</i> (1992) [33] by Pauline Oliveros • <i>Envelopes</i> (2020) [34] by Elizabeth Millar • <i>Two Definitions</i> (1963) [35] by George Brecht
Prose	• <i>February 26, 2000</i> (2000) [36] by Amnon Wolman • <i>Sextet: The Tigers Mind</i> (1967) [37] by Cornelius Cardew
Concrete	• <i>A Column of Sound</i> (2025) [38] by Sage Harlow • <i>eine/r 2</i>, (2008) [39] by Stefan Thut • <i>Piano Piano</i> (1988) [40] by Pauline Oliveros • <i>e.a.c.d.</i> (2014) [41] by Ryoko Akama

Table 2: categories of poetic text scores by word use and placement.

2.1 Single words or ideas

Scores utilising only single words, or single ideas, include Yoko Ono’s *Hide Piece* and *Cut Piece*, respectively stating only “hide” and “cut”; George Brecht’s *Suitcase* containing only the title and a bullet point with the word “suitcase”; *wind* and *grass*, by Lo Wie, scores existing only as titles transmitted orally by

² It is necessary to note that it is outside my means to create a comprehensive list of poetic text scores. Scores here have been included to balance historical precedent with select contemporary use.

the composer. My own piece *October 24, 2022* (Fig. 1) containing the words “Autumn” in addition to a duration.

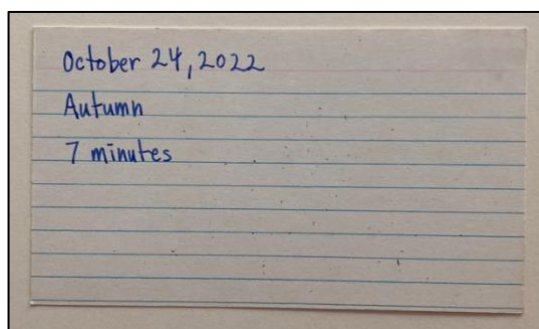


Figure 1: *October 24, 2022* by Craig Pedersen

2.2 Word groupings or lists

Examples of scores composed of groups of words or short lists include George Brecht’s *Three Aqueous Events* containing bullet points listing the three words “ice”, “water”, and “steam”; and Howard Skempton’s *for Strings*, composed of three handwritten words “waves”, “shingle” and “seagulls”. Of note are scores by Manfred Werder, each utilising a different approach to word grouping or lists. Examples include *2005*¹, using three words “place”, “time”, and “(sounds)”; *2008*³ (Fig. 2) a list in both German and English, composed of “eine anhöhe a hill / ein tal a valley / eine bergkette a mountain range / eine tiefebene a lowland / eine hochebene a plateau / ein flussdelta a river delta / ein fjord a fjord; *2012*(4), a list of three words in different languages, “litio” (lithium in Italian, Spanish, and Basque, from the Greek líthos, meaning stone), “φωνή” (the Greek word for voice, sound or noise), and “papier” (French for paper).

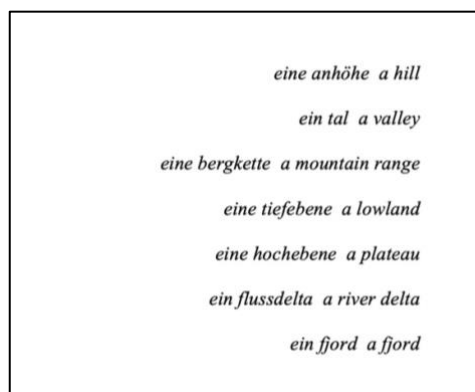


Figure 2: *2008*³ (detail) by Manfred Werder

2.3 Short statements, poetic or otherwise

Examples of pieces utilising short statements include George Brecht’s *Chair Event* composed of the poetic description “on a white chair / a Christmas -tree ball / flag / can opener / black and spectral colors”; La Monte Young’s *Piano Piece for David Tudor #3* reading “most of them / were very old grasshoppers”; Sarah Hughes’ *Sifted ashes of air*, containing a poetically arranged line of text attributed to Edward James, along with the ambiguous statement “For four / Letting most sounds fall

/ With some retained”; a series of scores of found quotes from literature or poetry composed by Manfred Werder; my own score *April 14, 2022* (Fig. 3) stating “asking to be part of the environment dismisses that we already are” in addition to duration, and *22 September 2024* [42] expressing “something will always remain, as a kind of residue, that exceeds analysis”.

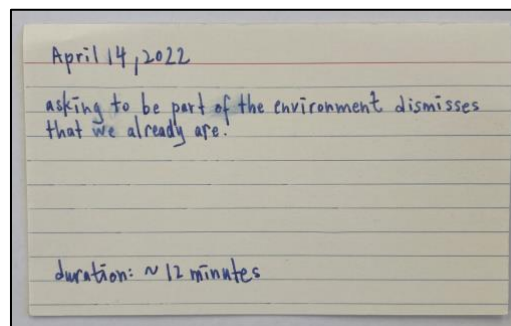


Figure 3: *April 14, 2022* by Craig Pedersen

2.4 Longer poetic verses

Although rarer, some poetic text scores exist as longer poetic verses. Examples include *The Earthworm Also Sings*, by Pauline Oliveros; *Two Definitions* by George Brecht, composed of two bullet points, one with a long ambiguous text, and another stating “(a cup and a saucer)”; *envelopes* (2022) (Fig. 4) by Elizabeth Millar, a long free-verse text describing qualities of sound.

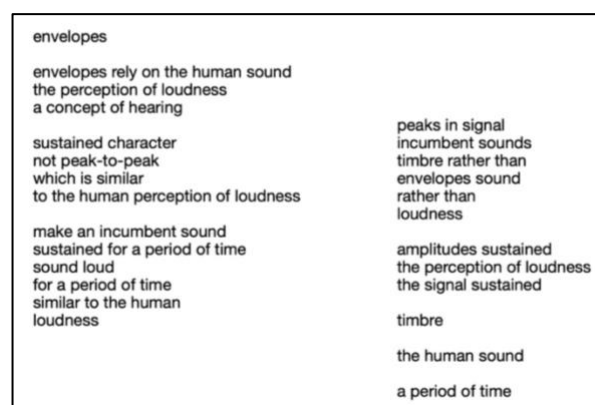


Figure 4: *envelopes* by Elizabeth Millar.

2.5 Long form prose

Rarer yet than scores composed of long-form poetic verses are those made of longer-form prose, such as Amnon Wolman’s *February 26, 2000*, a text describing an imaginary sounding experience:

A new sound, a very nervous, taut and nerve-racking sound. You hear it and almost want to put your finger in your ear to dampen it, but you hold back, reasoning that it will mutate, it will turn into something else, but it doesn’t. It is an annoying piercing shrill, similarly to a difference tone it is seemingly placed inside your ear. It persists and perseveres. It is not quite electronic, or not what we think of when we say an electronic sound [37, p. 419].

Another compelling example is Cornelius Cardew's *Sextet: The Tiger's Mind*. A hybrid instructional-poetic score, the work is separated into two parts. The first section contains two prose verses titled *Daypiece* (Fig. 5) and *Nightpiece*, revolving around six characters: Amy, the tiger, the tree, wind, the circle, and the mind.

Daypiece
The tiger fights the mind that loves the circle that traps the tiger.
The circle is perfect and outside time. The wind blows dust in tigers' eyes. Amy reflects, relaxes with her mind, which puts out buds (emulates the tree). Amy jumps through the circle and comforts the tiger. The tiger sleeps in the tree. High wind. Amy climbs the tree, which groans in the wind and succumbs. The tiger burns.

Figure 5: *Daypiece* (excerpt) from *Sextet: The tiger's mind* by Cornelius Cardew

The second section offers commentary on interpretation, in addition to defining features of each character. Of particular note is the way in which the second section offers no musical guidance. Instead, the commentary on interpretation discusses potential approaches to the work, such as performing *Daypiece* and *Nightpiece* on alternate occasions, memorisation of the text, assigning of roles, and discourse on ways in which the work could evolve. Further, while the descriptions of characters are akin to poetic characterisations, rather than musical motifs, as demonstrated below:

Amy is a person. She worships the tiger. She tags along holding him by the tail. Her mind is occupied with things close by. She comes to no harm in the wind, although it brings her intimations of things far away. However, in high winds she should avoid climbing trees [38, p. 530].

2.6 Concrete Poetry

In addition to the above mentioned word use, many scores use qualities of concrete poetry, where placement or arrangement on the page, in addition to change in typeface is also a poetic factor. Explicitly concrete works include *eine/r*, 2 by Stefan Thut, a score with single instances of the words “sound” and “noise” arranged on the page, and the words “once (again)” centre-justified in a smaller size at the bottom; *Piano Piano* by Pauline Oliveros, a score containing the words “play”, “black”, “white”, “one” and “tone” arranged in different placements, sizes and font-weights; *e.a.c.d.* by Ryoko Akama, composed of a time-line from 0’00” to 20’00” surrounded on the page by a number of words and short phrases, such as “small”, “holes in the eye” and “like a chimney”, including a longer text reading “silent letter is a letter that is not pronounced / yet without it / the word makes no sense”; *A column of sound* (fig. 6) by Sage Harlow, composed of the multiple instances of the phrase “a column of sound” overlaid in different ways on multiple pages.



Figure 6: *A column of sound* (excerpt) by Sage Harlow

It is noteworthy that regardless of category, most of the scores mentioned thus far take care with the spatial placement and visual presentation of words, in addition to material choice. For example, Brecht is known for use of dot points, with scores presented on small white or off-white cards; Werder's unmistakable use of right-justified text in the middle of a page, with lower-case signature and title at the bottom, often typewritten on found paper; the poetic placement of text in Sarah Hughes' *Sifted Ashes of Air*; my own scores, handwritten on a set of found index cards, written with the same pen and ink in a consistent format.

2.7 Categorisation by allusive, ambiguous, conceptual

The challenge of categorising scores by word use and placement lays in the lack of thematic connection between scores. For further thematic connection we turn to Anderson's definition of poetic text scores as allusive, conceptual, or ambiguous [4]. Although the terms are offered without further qualifiers, beyond that such scores consist of text which does not explicitly point to action, nor contain performance instructions, it is worthwhile to consider the viability of the terms as being categorically distinct. In this view, *allusive* scores primarily refer to the presence of an act, or a description of a result, without outright instructing; *ambiguous* scores are primarily ambiguous as pertains to their actionability; and *conceptual* scores primarily realised in the mind [44]. It is important to consider that while many scores fit within multiple categories, I have chosen to categorise according to what appears to be their primary characteristic, recognising this is an interpretive choice. In this light, previously mentioned scores can be categorised as follows in Table 3:

Allusive	• 2005¹ by Manfred Werder • October 24, 2022 by the Craig Pedersen • Chair Event by George Brecht • envelopes by Elizabeth Millar • Suitcase by George Brecht • Three Aqueous Events by George Brecht • wind and grass by Lo Wie
----------	---

Ambiguous	• 2008³, 2009¹, 2012⁴ by Manfred Werder • <i>A Column of Sound</i> by Sage J Harlow • <i>April 14, 2022</i> by Craig Pedersen • <i>for Strings</i> by Howard Skempton • <i>Piano Piece for David Tudor #3</i> by La Monte Young • <i>e.a.c.d.</i> by Ryoko Akama
Conceptual	• <i>Earth Piece</i> by Yoko Ono • <i>February 26, 2000</i> by Amnon Wolman • <i>The Earthworm Also Sings</i> by Pauline Oliveros • <i>Two Definitions</i> by George Brecht

Table 3: Categorisation of scores by allusive, ambiguous, and conceptual

As suggested, an allusive score primarily alludes to the presence of an act, or a description of a result, without outright instructing. In this light a score such as Werder's 2005¹ makes clear the necessary components for its realisation: place and time, with sound (written in parentheses), somehow acting as explanation or afterthought; Brecht's *Chair Event* takes meaning as the description of a potential realisation whereby a Christmas-tree ball, a flag, and a can opener are placed upon a white chair, with black and spectral colours present, while *Three Aqueous Events* describes three events, defined at minimum by the presence of ice, water, and steam, presumably in that sequence; my own score *October 24, 2022* ("Autumn") alludes to a realisation where aspects, qualities, documentation of, or responses to, the idea or archetypes of Autumn are present.

Conversely, ambiguous scores contain heightened degrees of ambiguity as pertains to their potential actionability. For example, Young's *Piano Piece for David Tudor #3* ("most of them / were very old grasshoppers") leaves no clear frame of reference for interpretation on the piano; Skempton's *for Strings*, presumably written to be performed on string instruments, leaves interpreters to consider the potential musical meaning of the words "waves", "shingle", and "seagulls"; Harlow's *A Column of Sound* presents an interpreter with the need to interrogate potential meanings of the phrase "a column of sound", in addition to negotiating the elements of concrete poetry discussed above. My own work *April 14, 2022* ("asking to be part of the environment dismisses that we already are") requires poetic unravelling, potentially including personal understanding of the philosophical meaning of the text and how it relates to praxis, to realise.

Drawing on the work of Doyle [43], we can consider conceptual scores as those primarily realised in the mind, whether through the presentation of impossible, impractical, or unthinkable to realise instructions, such as those found in many scores written by George Brecht, Yoko Ono or other Fluxus associated artists; inquiry or internal reflection scores such as many of Pauline Oliveros' sonic mediations; scores explicitly made to be realised in the mind, such as Wolman's *February 26, 2000*, the intended to function as guide to creating an imaginary sounding experience. Of interest is how Wolman's score can be viewed as quite unambiguous, even instructional, in its description of an imagined

music, but that its markedly conceptual nature cannot be overlooked.

In the above examples, qualities and characteristics of poetic text scores have been explored, and it has been demonstrated that they can take many forms. Further, although thematically varied by content, they can be loosely connected by their forms of ambiguities. However, while basic categorisation and simple description of their qualities and ambiguities is necessary to their understanding, I contend that to fully understand the nature of poetic text scores we must broadly consider what musical scores are and investigate the nature of poetic language. Doing so both provides context for understanding how poetic text scores function as scores and sets the foundation for understanding interpretation in relation to broader practice.

3. THE SCORE AS SEMIOTIC MEDIUM

In general, there are two main views of scores, that they are visual representations of music, or set of directions to a performer towards a musical event [44]. The mainstream view of music notation corresponds to the former, that scores and notation represent the way music sounds (or should sound), graphically or symbolically [44], [45], [46], [47], and it is generally accepted that notation serves a variety of purposes: performance and publication, to sketch and remember musical ideas, musical analysis, or to follow along with recordings [48]. Further, it has also been suggested that graphic representations of music, such as staff notation can be considered "descriptive", in that they describe the sound of the work (albeit with certain limitations and functions defined by style and genre), and notation which directs a performer in the method of making sound can be considered "prescriptive" [48]. This paper operates from the view that, at least to a performer, scores can be viewed as directions in that they signify acts in time, doing so not only descriptively or prescriptively, but also poetically. Further, regardless as to being viewed as visual representation of sound or as directions, all scores can be viewed as a system of symbols to be interpreted towards musical performance.

3.1 Overview of semiotics

Within semiotics we also recognise two schools, the "American" and "European", building upon the work of Pierce and Saussure respectively, through which the signs and symbols of music have been considered [49]. This paper draws primarily upon the work of the latter. In *Course in General Linguistics* [50], compiled posthumously in 1913 from lecture notes by his students, Saussure outlines a dyadic approach to understanding to linguistics signs, arguing two main components of any sign, which do not exist independently of one another: the *signifier* as the form or material vehicle of the sign, and the *signified* as concept, idea or meaning of the sign. He argues that language has both individual and social aspects which are inextricably linked, existing both as a product of the past as an established system, and in the present, synchronically. He further asserts that linguistic

signs do not exist independently of their context, which is a complex system, and that within different systems of meaning, signifiers signify differently, carefully noting that neither individuals nor the collective control the meaning of language. Although his work was fundamentally concerned with linguistics, he suggested the possibility to study signs more broadly, coining the term *semiology* (later known as semiotics). Following Saussure's work, theorist Roman Jakobson later pioneered the structuralist movement, broadly applying the linguistic model and insights into the nature of signs to other fields, such as philosophy, literary theory, anthropology and ultimately music.

3.2 Overview of score as semiotic medium

In her essay *Interlude I: Exploring musical integrity and experimentation*, Katherine Coessens [45] succinctly summarises the musical score in the terms of semiotics. Using the language of linguistics, she describes scores as semiotic mediums containing syntax and grammar, signification (representation of meaning) and semantics (meaning), and pragmatics (the way said language is used) in relation to performer practice. As such, she suggests that scores convey signification by way of signs, that scores are contingent upon the systems of notation defined by the rules and conventions of a given musical style, and that scores contain both indeterminate and seemingly fully determined elements, whose relationship to sounding outcome is contingent upon the interpreting musician's practice and their context. She argues (from a canonical Classical perspective) that although the "syntax of music notation is quite clear", having been developed over time by artists and communities through spoken and unspoken consensus, there is no verbatim way to render a score's notation (code or signs) into sound, thus suggesting a complex process of musical literacy, physical learning and translation required to perform a score. As such, she notes that while scores as signs convey meaning from one system to another, from score to sound, the sounding experience of the score is mediated through the body and mind - and thus practice - of the interpreter.

However, while Coessens suggests that musical scores "reduce musical ideas to symbols" whose interpretation are agreed upon, I contend that while this is something scores (and their surrounding context) *can* do, that her suggestion is a more apt description of only particular historical systems of notation. As we have seen, within Experimental music the sounding outcome of a score need not be known in advance, nor reproducible, nor do scores necessarily describe clear process to an interpreter. In this light, I suggest that scores can be read as systems of signs, and the signification of symbols into sounds are contingent upon the systems of notation (or lack thereof) defined by the rules and conventions of a given musical style and context, the sounding experience of which is mediated through the practice of the interpreter.

3.3 The score as object signifying acts in time

However, even this definition of the score presupposes an understanding of a score's fundamental identity, without which the system of signs has no context for an interpreter to understand said signification of symbols. As Pisaro [7] suggests, scores are "to be understood as a set of symbols whose goal³ is to be interpreted, in most cases, as sound (and silence)." In his suggestion that score *is to be understood* as something, Pisaro points to an implied fundamental identity signified by the score, without which the symbols therein lack interpretative significance. In short, musical symbols requiring an interpreter to realise must act as a form of instructions to a performer.

To this end, I propose considering the score as semiological chain, the first part of which establishes a score's fundamental identity: recognising the sign as both signifier and signified, we can broadly view the score as an *object signifying acts in time*. Once a score's identity is established as such, the second part of the semiological chain, understanding the signification of symbols vis-a-vis the act of interpretation, arises as answer to the questions "which acts, and when?"

What is powerful about this definition of the score, is that it holds true not only for different types of musical scores - standard notations, graphic notations, verbal text scores, conceptual text scores, and transcriptions, in addition to non-musical scores such as action scores, event scores, performance scores and dance scores. Further, viewing the definition *object* broadly, we can include aural or audio scores such as found in Gavin Bryars' score *1, 2, 1-2-3-4* [51] as sound object, or spoken scores such as Eliane Radigue's *Occam Ocean* [52] as speech object.

It bears repeating that this identity is not inherent to the identity of the score itself but is dictated by context. While on first observation this might appear to mean, as Anderson [4] suggests "a score (or notation or composition) is a score (or notation or composition) if the culture uses it as a score (or notation or composition)" (p. 136), as Saussure and later structuralists argued: we don't control the signified meaning of language and linguistic signs are ratified by collective understanding and agreement. Further, it is understood that within different systems of meaning, signifiers signify differently, and that context is a complex system. As such, it is a much wider context that comes into play to signify the score's observable definition of an *object signifying acts in time*, not its use. In this way, we can use a shoe as a score, but that doesn't make the shoe a score, it makes the shoe "used as a score".

3.4 Identifying the score

Further, it is necessary to consider another important factor preceding the semiological chain: the recognition of a score's identity. Considering that signs don't exist

³ In his definition of the score, Pisaro notes a "goal" of the symbols. It should be emphasised however, that symbols have no agency, only contextually defined understanding and use, and that said implied ascription of agency falls to the interpreter, in the process of interpretation.

independently of their context [51] and thus context dictates legibility as media [53] within a ‘semiosphere’ of signs [50], we view the aforementioned semiological chain and signification of symbols therein as illegible until we identify the score within a musical context. Therefore, the signs and symbols contained within a musical score, once removed from the purview of Western Classical music appear unintelligible. Given a non-musician will inevitably experience a musical sign differently than a musician [54], the score could easily be misidentified a treasure map or cipher. Said misidentification however would simply be speculative, especially given that the score already has a defined identity in another context. However, said defined identity doesn’t preclude the possibility of a musical score being a treasure map in another context. Hypothetically, within some fantastically arranged series of events a score and its symbols could be understood to be, and used as, a treasure map.⁴

Even within a musical context, some scores are more readily identifiable as scores than others. It will always be easier to recognise a musical score using the syntax of music notation situated well within the limits of tradition, than those who expand, move away from or challenge said limits, due to the familiarity borne from the marriage of historical precedent and education supporting musical literacy. As noted, musical syntax has been developed over time by artists and communities though spoken and unspoken consensus. To this end, we also understand why theories about music (including my own) may lose their authority as we move away from them in time: language and its systems exist both as a product of the past as an established system, and in the present, synchronically, constantly changing, outside of both individual and collective control. In short, scores that move away from the symbolic precedents of the past require a new or modified discourse that accounts for them.

3.5 Other semiological approaches

Accordingly, viewing the music semiotically isn’t novel, having been explored in detail in Nattiez’ [55] *Music and Discourse: Toward a Semiology of Music*, and Goodman’s [56] *Languages of Art*, with a clear history found in Tarasti’s [50] *Signs of Music: A Guide to Musical Semiotics*. However, Tarasti’s emphasis on historical canonical works, musical meaning of a score’s sounding outcome, or notational relationships is not relevant to my research. Further, given my research deals with scores synchronically, as things already made and presented to an artist for musical or sounding interpretation, it does not deal with composer process. It intentionally eschews notions of hidden composer intention carrying through the score to the interpreter, taking the stance that composer intention is fundamentally unknowable, perhaps even to composers themselves, and is best considered a form of context. Thus, Nattiez’ emphasis on the poetic - the intended meaning or process of creation is also not relevant.

⁴ Cue Dan Brown or Nicolas Cage.

Similarly, Goodman’s theory too rigidly defines the score as a system of symbols containing fixed meanings, discounting Experimental graphic or text scores on the basis of being too non-notational [4], [57]. Further, his theories have been widely criticised within classical canon as insufficiently accounting for the way a score’s use, and thus understanding and relevances of symbols therein, is dependent upon context [57].

Finally, each of the proposed theories above sets out a series of complex systems, none of which are particularly useful to a musical interpreter, nor provides a concise way to view the interpretation of poetic text scores (or any scores for that matter). They mostly operate from the vantage point that the symbols presented in a score are understood to represent the sounding end result of a process, and that the codes of notation (or non-notation) are readily understood. In short, they leave out the essential perspective and process of the interpreter, without which the signs are not actualised. Thus, I return to the simple semiotic model proposed by Saussure of the sign and the signified, the definition of the score as object signifying acts in time, and understanding that the signification of symbols vis-a-vis the act of interpretation arises as answer to the questions “which acts, and when?”

3.5 Instructional text scores

Although instructional text scores are not the primary focus of this paper, the ways in which they descriptively or prescriptively indicate to an interpreter “which acts, and when” demonstrates the ways in which text signifies acts in time, providing context and highlighting the fundamental ambiguities present poetic text scores. The examples here are chosen because they quite elegantly and clearly demonstrate some of the ways scores signify acts in time but are by no means comprehensive to the multitude of possibilities and combinations.

Antoine Beuger’s *one tone. rather short. very quiet.* (1998) [58] (fig. 7) is a beautiful example of a text score describing a sounding result, albeit within certain accepted boundaries of knowability present in open scores.

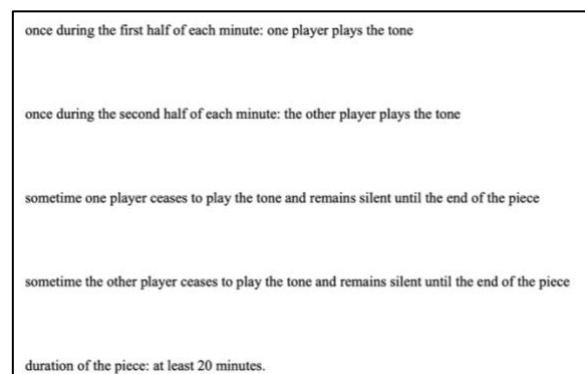


Figure 7: *one tone. rather short. very quiet.* (excerpt) by Antoine Beuger

As Saunders [59] elegantly summarises:

Players each make a sound which is rather short and very quiet. One player makes the sound in the first half of every minute, and the second player makes the sound

in the second half of every minute. After some time, each performer then independently stops making the sound and remains silent until the end of the piece (p. 106).

In this way, the score clearly describes the act of playing one short single tone in the first or second half of every minute, until a performer chooses to stop. Although the score may not describe a specific pitch or tone or its discrete placement in time, it nonetheless indicates moderately clearly what to do and when.

James Tenney's swell pieces (fig. 8), from the composer's series of *Postal pieces* (1971) [60], are another set of text scores clearly describing musical acts. In *Swell Piece: for Alison Knowles* (1968) [61] a group of three or more performers are asked to play long tones, one after another, each tone building from as "softly as possible" to "maximum intensity" before fading into silence, while keeping the pitch and timbre steady. However, the specific temporal placement of each tone within the agreed upon duration of the piece is not described. Instead, placement in time is prescribed through the process of playing each tone sequentially, as dictated by each performer's musical choices, instrument, and capabilities, no doubt in addition to a multitude of other factors. Both *Swell Piece No. 2* (1971) [62] and *Swell Piece No. 3* (1971) [63] follow in a similar fashion; performers are told what to play, swells on A-440 and B3/F#4 respectively, while placement in time is prescribed by a process. In both cases performers play independently of one another, leading generatively to a series of overlapping swells, temporal aspects again dictated by performer's musical choices, instrument, capabilities, and other factors.

Figure 8: Swell piece: For Alison Knowles, Swell piece no. 2, Swell piece no. 3 by James Tenney

While Steve Reich's *Pendulum Music* (1968) [64] (Fig. 9) is primarily prescriptive, it also contains descriptive elements. The score prescribes the procedure of suspending microphones by their cables such that they can swing in a pendular motion above loudspeakers to create feedback, stating the process by which the microphones are set in motion and ultimately come to rest. While the score describes the sounding result as a "series of feedback pulses...which will either be all in unison or not", such outcome is fundamentally dependent upon the relationship between each microphone-

pendulum and loudspeaker relative to the changes in each pendulum's oscillation period.

Figure 9: Steve Reich's *Pendulum Music*

The above scores demonstrate ways in which instructional text scores can signify acts in time, both descriptively and prescriptively. However, some instructional scores are more ambiguous, broad, or open-ended in the ways they signify, at times neither describing or prescribing what to do, or when.

Sage J Harlow's *Thirteen pages of improvisation* (2017) [65] (Fig. 10) is primarily prescriptive but contains implied or ambiguous elements. The cover page indicates that it is a piece for one or more improvisors, with open instrumentation, prescribing that interpreters play from the score and not from memory, additionally designating a minimum duration. Each of the subsequent thirteen pages contains only two elements: the word *improvise* slightly above the midline of the page, and the page number in the bottom right corner. Unlike previously discussed scores, the sounding result is in no way described (unless of course you view improvisation as being a specific sound, a notion I would argue against). Instead, the score prescribes that the interpreter *improvise*. Further, while the score tells an interpreter which act to perform - improvisation - it does not designate the instruction's placement in time, other than occurring within the stated minimum duration. Although the use of a single imperative verb is most readily read as a command, implying placement in time as "now" or "within the context of performing this piece", the score's ambiguous nature invites multiple interpretations, without clearly marking their limits. For instance, an interpreter could hypothetically improvise for a period proportional to the space the word occupies on the page. Considering this, we might imagine the interpreter's role not only as the interpretation of prescriptive or descriptive symbols present in the aforementioned second part of the semiotic chain (the set of symbols to be interpreted as sound and silence), but to engage in an act of interpretation beginning with the first step of the chain (the ways in which it signifies acts in time). In the case of *Thirteen pages of improvisation*, as act of interpretation the performer decides upon or takes a position (implicitly or explicitly) regarding the way the score signifies time and

interprets the score accordingly. In this way, lacking prescription or description to guide them, an interpreter engages with the nature of the question “when?”, setting the foundation for which they then realise the symbols as presented. Noteworthy is how despite the ambiguity present in the *Thirteen pages of improvisation*, the score still instructs the performer and is thus categorically an instructional text score.

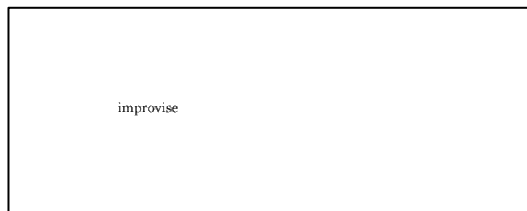


Figure 10: Sage Harlow’s *Thirteen pages of improvisation* (excerpt)

4. (RE-)DEFINING THE POETIC SCORE

Of central interest to this paper are text scores which neither clearly describe or prescribe acts *nor* placement in time. As noted above, my preference is for the terms poetic text score as such scores are defined by their poetic nature. As Montgomery suggests:

The poem is the utterance that most tests the meaning-making capacity of language. In the parallel practice of the poem, the text score finds its most potent analogue, in which the lucid instruction for performers dissolves under the pressure of entirely different conventions of reading and writing. [3, p. 90]

Montgomery further notes, through analysis of composite instructional and poetic scores by Michael Pisaro and more minimal non-instructional text scores of Manfred Werder, that while use of poetry and poetic language in scores does not provide clear purpose to an interpreter, it may serve to orient a performer. He discusses how within Pisaro’s works poetic qualities set conditions for an interpreter to respond to other forms of notation present in a work, and thus bring sounding qualities to a work that would be otherwise absent, while Werder’s scores offer vectors of exploration in and out of the work, acting as a “frame for the world” [3, p. 99]. However, while Montgomery’s observation of the use of poetic language is apt, of most interest is his suggestion that text can “exert a linguistic pressure on the performative realisation of the poem.” [3, p. 93] As follows I suggest that it is not the use of poetic language or poetry nor the implementation of para-poetic qualities alone that define such scores as poetic, but that they function poetically, in linguistic terms.

4.1 The poetic function

To better understand the linguistic properties of poetry, and thus poetic text scores, we turn to the work of Roman

Jakobson.⁵ In his *Closing Statement: Linguistics and Poetics* Jakobson [15] outlines six factors involved in verbal communication - addresser, addressee, context, message, contact, and code:

The ADDRESSER sends a MESSAGE to the ADDRESSEE. To be operative the message requires a CONTEXT...; a CODE fully, or at least partially, common to the addresser and the addressee; and, finally, a CONTACT, a physical channel and psychological connection between the addresser and addressee, enabling both of them to enter and stay in communication (p. 353)

Linda R. Waugh [66, p. 58] expands on these factors as follows:

1. addresser (speaker, [...] poet, author; narrator)
2. addressee (decoder, hearer, listener; reader; interpreter)
3. code (system, langue⁶)
4. message ([...] parole, the given discourse, the text)
5. context (referent)
6. contact

It is relatively straightforward to apply this model to the context of musical score interpretation: The *addresser* can be viewed as either the composer, or the literary voice of the score itself⁷; the *addressee* the interpreter; the *code* the system of symbols and signs - traditionally music notation and its syntax; the *message* the text itself, or the specifically used musical symbols; the *context* both the specific situation surrounding a given work, in addition to rules and conventions of a given musical style and context, and interpreter practice; *contact* the way the score is in dialogue with an interpreter’s practice and genre as means to understand and realise the score, or if taking the addresser as composer (or other musical mediator such as a music director) the channel of communication. It should be noted that these same factors could apply between interpreter/sounding of the score and listener, but that this lies outside of the scope of this study.

Jakobson further suggests six functions of language, each corresponding to one of the six factors:

1. emotive, focusing on language the addresser uses to express their internal state;
2. conative, directly communicating with the addressee, primarily using imperative (commands) and vocatives;
3. metalingual, communication used to clarify the code;
4. poetic, language which deals with the message or form itself;
5. referential, descriptive statements, including those referring to time, place or person;
6. phatic, messages pertaining to the establishment, continuation or ending of communication.

Jakobson quite importantly notes that each function does not exist in a vacuum, and that it is rare that verbal messages only fulfil a single function. As Waugh suggests, verbal messages are multifunctional, each

⁵ As McKay argues in *A Semiotic Approach to Open Notations* [54], Jakobson’s work is essential to understanding open notations, and thus poetic text scores.

⁶ *Langue* referring here to the shared system of language, and *parole* the individual act of using language, as defined by Saussure.

⁷ Additionally, within post-structuralism, in absence of a present composer, we take the score to be the addresser.

fulfilling diverse functions integrated within a single message, but depending “primarily on the predominant function” [15, p. 353]. Although also outside the main focus of this paper it is not difficult to conceive of compositional practice as relating to Jakobson’s six functions, at least when viewed from the perspective of the composer. For example, some work is primarily *emotive* in that aims to express a composer’s internal or emotional state, other *referential* in the sense it is written “about something”, while some practice is centred on the limits or potentials inherent in musical *code*.

Although Jakobson defined six functions of language, his central interest lay predominately with the poetic function, and the ways language and linguistic form constitute the aesthetic centre of the poetic genre. In Jakobson’s view the poetic function focuses primarily on the intrinsic qualities of the message or form itself, with meaning supporting the form rather than the form supporting meaning, thus revealing linguistic form as being essentially human and exposing fundamental subjectivities of language [67]. As Waugh (1980) suggests, “the set toward the message brings forth the question of the nature of the message” (p. 60).⁸ Viewed in the language of sign and signified, the signified illuminates the nature of the sign.

As we have already demonstrated with Harlow’s *Thirteen pages of improvisation*, the use of ambiguous language in instructional scores may impel an interpreter to engage with the nature of the score and make interpretative decisions (implicitly or explicitly) about the ways in which the score signifies time. It is not a large leap to consider that scores completely lacking in instructional function similarly. Thus, without the expected instructional symbols that otherwise define a score, the interpreter of scores lacking in instruction engages with the nature of the form itself, and interpretation will inevitably be based on an interpreter’s understanding of what a score is and how the symbols or code presented do or do not relate to said understanding. Further, depending on interpreter experience they may engage metalingually, using language to clarify the nature of the score, asking the questions “how is this a score and what do I do?”. Indeed, as previously noted, ambiguity is a feature of poetry, and poetic text scores as are in part defined by the ways in which their ambiguity extends to their meaning, definition and interpretability as scores.

As with the poetic function, the meaning of poetic text scores supports its form. As a form of open score featuring a heightened degrees of sounding difference between performances, a non-instructional text score’s is defined by how it is used and how interpretations existing within a larger field relate to one another and point back to the original text. As Montgomery elegantly states “travelling in the opposite direction, the performance of the score ... always gestures back towards the words that give rise to it and which condition the sounds in ways that apply in no other compositional or improvisational domain” [3, p. 87]. As such the score is not defined by how interpretations function as reproductions of an

original score,⁹ but instead in how they function as manifestations contributing to the expressive potential of a greater work. In short, interpretations tell us what the score is, as opposed to the other way around.

Further, in the same ways poetry reveals that linguistic form as being essentially human and expose fundamental subjectivities of language, poetic scores analogously by nature expose the human aspects of score interpretation. It is not difficult to conceive of how the sounding end-result and qualities thereof of a given realisation of a score is shaped by the skills, abilities, and aesthetic practice of the interpreter. As such, poetic scores highlight the subjective aspect of score interpretation, re-emphasising interpreter practice and humanity present in the interpretation of all works.

5. EPILOGUE: FURTHER MUSING ON GENRES

Expanding on Jakobson’s work, Cureton [67] suggests the ways that different literary genres use language in distinct manners, consequently leading to different aesthetic centres for each genre. For example, he posits the predominant use of language within drama as sociolinguistic and pragmatic, centring aesthetically on the ways language is used in different contexts; prose as semantic and rhetorical, centring on ideas and what they point to; song as paralinguistic and physical, centring on the ways we use factors such as pitch, volume, and intonation to modify meaning. As with Jakobson and Waugh, Cureton carefully notes that while these are observable predominant features of each genre, they additionally employ features from other genres.

Viewing text scores as a hybrid literary-musical genre we can similarly consider the distinct ways in which both instructional and poetic text scores use language, and thus their aesthetic centres. By Jakobson’s categorisation instructional text scores are conative, as the text directly communicates to the interpreter using primarily directive or imperative language, descriptively or prescriptively, to clearly communicate actionable musical ideas, as demonstrated in the previous section. As follows, the aesthetic centre of instructional text scores is not unlike song, being both physical and musical, focusing on ways in which acts in time are incited towards musical ends, including factors like the paralinguistic - pitch, volume, intonation, rhythm, timing, and timbre.¹⁰ To an interpreter, instructional text scores provide relatively clear pathways towards musical realisation.

Conversely the central use of language in poetic text scores is decidedly non-conative, lacking directive or imperative language, instead commonly considered poetic by nature of their suggestive or enigmatic qualities

⁸ For further detailed linguistic explanation, see Waugh’s *The Poetic Function in the Theory of Roman Jakobson* [66].

⁹ This parallels Susan Sontag’s, John Berger’s and Walter Benjamin’s respective views on photography, art, and translation.

¹⁰ Although paralinguistic and musical expression share many qualities, the distinction is made here to account for the field of paralinguistics’ primary focus on *vocal* utterances, and additional comparisons between paralinguistic and musical expression lie outside this work.

and use of ambiguously language. As hybrid literary-musical genre we understand their aesthetic centre to be poetic for the reasons demonstrated above, and musical¹¹, in that expression is towards musical ends. However, given a score signifies acts in time, and the poetic text score's aesthetic centre must also be physical. As such an interpreter is presented with an apparent paradox: how does a poetic text score signify acts in time while the language therein provides relatively unclear pathways towards musical realisation?

The simple answer is that poetic text scores signify acts in time poetically, and as such interpretation is a poetic process, involving engagement with both form and content of a score, shaped by the life and practice of the interpreter. Notably, Jakobson also viewed the effects of poetic language as subliminal and subconscious, suggesting the possibility to consciously articulate the structure and effect of poetry as secondary to its pre-lingual effects. While definitively proving that poetic scores signify to an interpreter subliminally, subconsciously, or pre-lingually lies elsewhere, it is beautiful to consider the clear parallel to be drawn here in the interpretation of poetic text scores, projecting a potential for interpretation of scores to lay outside of fully controlled or rational processes.

6. REFERENCES

- [1] J. Lely and J. Saunders, *Word Events: Perspectives on Verbal Notation*. New York: Continuum, 2012.
- [2] J. Walshe, *The Text Score dataset 1.0*. Darmstädter Ferienkurse, 2021.
- [3] W. Montgomery, "Text-Score-Text," in *Writing the Field Recording*, S. Benson and W. Montgomery, Eds., Edinburgh University Press, 2018, pp. 86–105.
- [4] V. Anderson, "The beginning of happiness: Approaching scores in graphic and text notation," in *Sound & Score: Essays on Sound, Score and Notation*, P. de Assis, W. Brooks, and K. Coessens, Eds., in Orpheus Institute Series. Leuven: Leuven University Press, 2013, pp. 130–142.
- [5] G. Bryars, "On verbal notation," in *Word Events: Perspectives on Verbal Notation*, J. Lely and J. Saunders, Eds., Continuum, 2012, pp. 124–126.
- [6] A. Lucier, "Notes on verbal notation," in *Word Events: Perspectives on Verbal Notation*, J. Lely and J. Saunders, Eds., Continuum, 2012, pp. 253–254.
- [7] M. Pisaro, "Writing, Music," in *The Ashgate Research Companion to Experimental music*, J. Saunders, Ed., Routledge, 2016, pp. 27–76.
- [8] C. Cox and D. Warner, *Audio Culture: Readings in Modern Music*. Continuum, 2004.
- [9] J. Gottschalk, *Experimental Music Since 1970*. Bloomsbury, 2016.
- [10] J. Saunders, *The Ashgate Research Companion to Experimental Music*, 1st ed. Routledge, 2009.
- [11] J. Cage and A. Knowles, *Notations*. Something Else Press, 1969.
- [12] T. Sauer, *Notations 21*, 1st ed. New York: Mark Batty Publisher, 2009.
- [13] L. Young, "Piano piece #3 for David Tudor (1960)," *An Anthology of Chance Operations*. New York Fluxus, 1963.
- [14] G. Brecht, *Water Yam*. Parrot Impressions, 1972.
- [15] R. Jakobson, "Closing Statement: Linguistics and Poetics," in *Style in Language*, T. A. Sebeok, Ed., in M.I.T. paperback series. , M.I.T. Press, 1964, pp. 350–377.
- [16] S. P. Su, *Lexical Ambiguity in Poetry*. London: Longman, 1994.
- [17] C. Pedersen, "October 24, 2022." *Mystery & Wonder*, 2022.
- [18] Y. Ono, "Hide Piece (1961)," *Grapefruit: A Book of Instruction and Drawings*. Simon & Schuster, 2000.
- [19] Y. Ono, "Cut Piece (1964)," *Grapefruit: A Book of Instruction and Drawings*. Simon & Schuster, 2000.
- [20] G. Brecht, "Suitcase (1963)," *Water Yam*. Parrot Impressions, 1972.
- [21] L. Wie, *wind*. self-published, 2007
- [22] L. Wie, *grass*. self-published, 2007
- [23] M. Werder, *2005¹*. Edition Wandelweiser, 2005.
- [24] M. Werder, *2008³*. Edition Wandelweiser, 2008.
- [25] M. Werder, *2012⁴*. Edition Wandelweiser, 2012.
- [26] H. Skempton, "for Strings (1969)," *Word Events: Perspectives on Verbal Notation*. Continuum, London ; New York, p. 350, 2012.
- [27] G. Brecht, "Three Aqueous Events (1961)," *Water Yam*. Parrot Impressions, 1972.
- [28] M. Werder, "2010¹." Edition Wandelweiser, 2010.
- [29] G. Brecht, "Chair Event (1963)," *Water Yam*. Parrot Impressions, 1972.
- [30] Y. Ono, "Earth Piece (1963)," *Grapefruit: A Book of Instruction and Drawings*. Simon & Schuster, 2000.
- [31] C. Pedersen, *April 14, 2022*. *Mystery & Wonder*, 2022.
- [32] S. Hughes, *Sifted Ashes of Air*. self-published, 2022. Accessed: Jan. 02, 2026. [Online]. Available: <https://sarahhughes.info/Scores>
- [33] P. Oliveros, "The Earthworm Also Sings (1992)," *Anthology of text scores*, S. Golter and L. Hall, Eds., Deep Listening Publications, 2013, pp. 1–16.

¹¹ Non-musical scores would have a non-musical aesthetic centre.

- [34] E. Millar, *envelopes*. self-published, 2020.
- [35] G. Brecht, "Two Definitions (1963)," *Water Yam*. Parrot Impressions, 1972.
- [36] W. Amnon, "February 26, 2000," *Word Events: Perspectives on Verbal Notation*. Continuum, London ; New York, p. 419, 2012.
- [37] C. Cardew, "Sextet: The Tiger's Mind," *Music Times*, vol. 108, no. 1492, p. 527, June 1967, doi: 10.2307/951366.
- [38] S. Harlow, *a column of sound*. self-published, 2025. Accessed: Jan. 02, 2026. [Online]. Available: <https://www.sagepbbbt.com/wp-content/uploads/2025/11/a-column-of-sound.pdf>
- [39] S. Thut, *eine/r*, 2. Editions Wandelweiser, 2008.
- [40] P. Oliveros, "Piano Piano (1988)," *Anthology of text scores*. Deep Listening Publications, p. 19, 2013.
- [41] R. Akama, *e.a.c.d.* self-published, 2014.
- [42] Craig Pedersen, *22 September 2024*. Mystery & Wonder, 2024.
- [43] K. C. Doyle, "Exquisite mind, euphoric knowing: Sound, language, and consciousness in experimental music and conceptual art after 1960," Case Western Reserve University, 2018.
- [44] C. Fox, "Opening offer or contractual obligation? On the prescriptive function of notation in music today," *Tempo*, vol. 68, no. 269, pp. 6–19, July 2014, doi: 10.1017/S0040298214000023.
- [45] K. Coessens, "Interlude I: Exploring musical integrity and experimentation," in *Sound & Score: Essays on Sound, Score and Notation*, P. de Assis, W. Brooks, and K. Coessens, Eds., in Orpheus Institute Series. Leuven: Leuven University Press, 2013, pp. 5–11.
- [46] P. de Assis, "Prelude," in *Sound & Score: Essays on Sound, Score and Notation*, P. de Assis, W. Brooks, and K. Coessens, Eds., in Orpheus Institute Series. Leuven: Leuven University Press, 2013, pp. 5–11.
- [47] C. Seeger, "Prescriptive and descriptive music-writing," *Music. Q.*, vol. XLIV, no. 2, pp. 184–195, 1958, doi: 10.1093/mq/XLIV.2.184.
- [48] M. Kanno, "Prescriptive notation: Limits and challenges," *Contemp. Music Rev.*, vol. 26, no. 2, pp. 231–254, Apr. 2007, doi: 10.1080/07494460701250890.
- [49] E. Tarasti, *Signs of Music: A Guide to Musical Semiotics*. Berlin: De Gruyter, 2012. doi: 10.1515/9783110899870.
- [50] F. de Saussure, *Course in General Linguistics*, London, England: Bloomsbury, 2013. doi: 10.5040/9781350284968.
- [51] G. Bryars, "1, 2, 1-2-3-4," *Verbal Anthology*. Experimental Music Catalogue.
- [52] J. Eckhardt, "Eliane Radigue OCCAM OCEAN," *Interpretations. – the book*, q-o2, 2014, pp. 74–79.
- [53] C. D. Dworkin, *No Medium*. Cambridge, Massachusetts: The MIT Press, 2013.
- [54] T. McKay, *A Semiotic Approach to Open Notations: Ambiguity as Opportunity*, 1st ed. Cambridge University Press, 2021. doi: 10.1017/9781108884389.
- [55] J.-J. Nattiez, *Music and Discourse: Toward a Semiology of Music*. Princeton, N.J: Princeton University Press, 1990.
- [56] N. Goodman, *Languages of Art*, Bobbs-Merrill, 1968.
- [57] L. Goehr, *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Oxford University Press, 1994. doi: 10.1093/0198235410.001.0001.
- [58] A. Beuger, *one tone. rather short. very quiet*. Edition Wandelweiser, 1998.
- [59] J. Saunders, "Commentary: one tone, rather short. very quiet," in *Word events: Perspectives on Verbal Notation*, J. Lely and J. Saunders, Eds., New York: Continuum, 2012, pp. 106–108.
- [60] J. Tenney, *Postal Pieces*. Smith Publications, 1971.
- [61] J. Tenney, "Swell Piece: for Alison Knowles," *Postal pieces*. Smith Publications, 1971.
- [62] J. Tenney, "Swell Piece No. 2," *Postal Pieces*. Smith Publications, 1971.
- [63] J. Tenney, "Swell Piece No. 3," *Postal pieces*. Smith Publications, 1971.
- [64] S. Reich, *Writings on music: 1965-2000*. Oxford University Press, 2002.
- [65] S. Harlow, *Thirteen pages of improvisation*. self-published, 2017. Accessed: Jan. 06, 2026. [Online]. Available: <https://www.sagepbbbt.com/scores/>
- [66] L. R. Waugh, "The Poetic Function in the Theory of Roman Jakobson," *Poet. Today*, vol. 2, no. 1a, p. 57, 1980, doi: 10.2307/1772352.
- [67] R. D. Cureton, "Jakobson Revisited: Poetics, Subjectivity, and Temporality," *J. Engl. Linguist.*, vol. 28, no. 4, pp. 354–392, Dec. 2000, doi: 10.1177/00754240022005108.

Acknowledgements

This research was supported by the Commonwealth through an Australian Government Research Training Program Scholarship.

[DOI: <https://doi.org/10.82133/C42F-K220>]

SOUND AS SCORE IN ARCHITECTURAL PLANS: A PATH-BASED NOTATION WORKFLOW USING AMBISONICS + 360° VIDEO

DI Nicolaj Kirisits

University of applied arts Vienna
Institute fine arts & media art
Department digital arts
nicolaj.kirisits@uni-ak.ac.at

ABSTRACT

Architecture is experienced in physical space but is represented as a non-temporal object in 2D plans, whereas sound is framed as a temporal phenomenon in the time–frequency domain. Bridging this divide requires conceiving architecture as event and sound as a spatial, architectural element. We propose a plan-centric notation that “folds” listening events along a walked path into the plan for re-enactment, comparison, and design discussion. Building on the *Soundspace-Sequenz*—an improvisational listening walk documented for later analysis—our method combines first-order Ambisonics (FOA) audio with 360° video. Python-based processing extracts descriptors (amplitude envelope, spectral complexity, tonal segments, onsets) and maps time to walked distance using video-assisted position checks. The resulting path-line is registered to the route and embedded in the plan; interpretive icon annotations (atmospheres, sound effects, soundmarks) are added manually under an explicit rule set. A case study in an urban public space shows how the plan-embedded score functions as a memory-stable record of listening and an actionable artifact for communicating sonic design concerns. Unlike static sound maps that merely locate sound, the notation documents listener–space interaction as constitutive of sonic reception.

INTRODUCTION

Architectural design workflows largely operate through 2D representations of physical space (plans and sections), whereas sound is commonly treated as an engineering constraint or reduced to aggregated soundscape indicators. While such indicators effectively describe larger urban spaces (macro: 10–100 m; macro+: >100 m), they often miss meso- and micro-scale phenomena crucial to architectural perception: thresholds, courtyards, passageways, edges, and the rhythmic coupling between locomotion and shifting audible situations (meso: 0.1–10 m; micro: <0.1 m).

This paper introduces a workflow that makes such situations legible in architectural plans—not by rendering sound as a static architectural element, as in many

soundscape representations [1] and early sound-art visualizations [2] [3], but as plan-embedded scores of situated listening. The contribution is a plan-native, derived score that couples perceptual events to route geometry through time-to-distance transfer, enabling repeatable reading and comparison within architectural drawing conventions.

The plan thus becomes a medium in which time is folded into space: a listening performance is recorded and converted into a path-based representation, where descriptor layers and annotated events are arranged along the traversal. The method is motivated by three observations: (1) architectural listening is enacted, not merely observed; (2) a traversal is both time and distance and can therefore be scaled and drawn; and (3) if a plan shows when and where sonic features occur along a path, the drawing can operate like an audio score for re-listening, comparison, and design discussion.

Beyond documenting “where sound is,” the goal is to represent how sound is co-produced through movement, orientation, and architectural affordances such as edges, openings, reflective surfaces, and thresholds. The plan is therefore treated as a coordinate system for a listening performance: changes in pace, stopping, and turning are part of the sonic situation and should remain readable in the notation.

We ask: (i) How can a spectral-temporal description of sound be translated into a plan-native representation without losing sequential and embodied aspects of listening? (ii) What minimal descriptor set remains legible for designers while enabling comparison across traversals? (iii) How can the output function both as documentation and as a re-performable audio score?

BACKGROUND AND RELATED WORK

From soundscape maps to scored traversals

Soundscape mapping and standardized vocabularies offer valuable frameworks for describing acoustic environments [4], but typical outputs (contours, indices, averages) can downplay sequential, embodied listening. For architectural purposes the challenge is to represent not only where

certain levels occur but how sonic situations unfold while listeners interact with space by moving, stopping, accelerating, or turning.

Soundwalk and listening-walk practices address this gap if they treat listening as situated action rather than as passive exposure. However, their outputs often remain narrative (field notes) [5] or media-based (recordings) [6] and are difficult to integrate into plan-centered design workflows. A plan-embedded, path-based score keeps route choice, speed changes, and thresholds visible, supporting side-by-side comparison between repeated traversals and making the interaction between listener and built form part of the representation.

Notation, plans, and folding

The technique of the *revolved section* (German: *Klappschnitt*) inserts sectional information into plans to clarify spatial change. Here a similar logic is applied: not a building section, but a perceptual event sequence is folded into the plan, preserving the plan as a shared design medium while adding a score-like time/distance dimension.

Folding is not merely a graphic trick but a way to establish commensurability between domains. Here, plan coordinates (meters) are connected to the spectral-temporal domain (seconds and descriptor values) via the walked path. This preserves the plan as a shared medium while avoiding the illusion that sound can be captured as a static layer independent of traversal and attention.

Soundspace-Sequenz: listening as architectural method

The workflow is grounded in the concept of the *Soundspace-Sequenz*: an improvisational traversal guided by *architectural listening*—an embodied, situated mode of attention that renders sound-based spatial elements legible as part of architectural form—and understood as a compositional act within a site. A single sequence is a singular event; a space can be considered a topology of many possible sequences. From such performances, a derived score can be generated and returned to representation as a re-performable artifact.

Audio scores, listening guides, and derived notation

Within expanded notions of scores—including audio scores, listening guides, and representational formats that organize attention—a score can mediate a performance even when the primary activity is listening. From this perspective, the *Soundspace-Sequenz* aligns with practices that treat listening as performative and generative: it produces a structured sequence of perceptual events that can be made amenable to notation. Related work also points toward derived forms of scoring, where notations are generated from recorded traversals rather than authored in advance, yielding re-enactable artifacts. Situated within architectural representation, such derived scores connect performative, time-based experience to plan-based conventions, supporting critique and iteration while

keeping rhythmic and atmospheric change legible along movement through space.

From visual path notation to audio-based scoring

Historical precedents for path-based notation can be found in architectural and urban design traditions that describe perception in sequence, such as serial-vision and route-based analyses (e.g., townscape approaches and “view-from-the-road” notations) [7], as well as practices that frame movement through space as a guided promenade. [8] Related lineages also emerge outside architecture: dance choreographers and notation pioneers developed systems to score movement and timing—codifying trajectories, pauses, and rhythmic transitions of bodies in space. [9] Across these precedents, the path becomes a representational device by registering changes in views, landmarks, edges, thresholds, and experiential rhythm as one moves. Yet what was typically “scored” remained overwhelmingly visual (or purely kinematic): they documented what can be seen, and sometimes how bodies move, while sonic phenomena—atmospheres, sound effects, and audio-based events co-produced through listener-space interaction—were largely absent or relegated to anecdotal description. The present work extends path notation beyond visual sequencing and movement coding by making sound-based spatial elements part of the ontology of architecture and notable within plan conventions, so that interaction with the built environment is documented not only as changing sightlines and motion, but also as changing listening conditions.

METHOD OVERVIEW

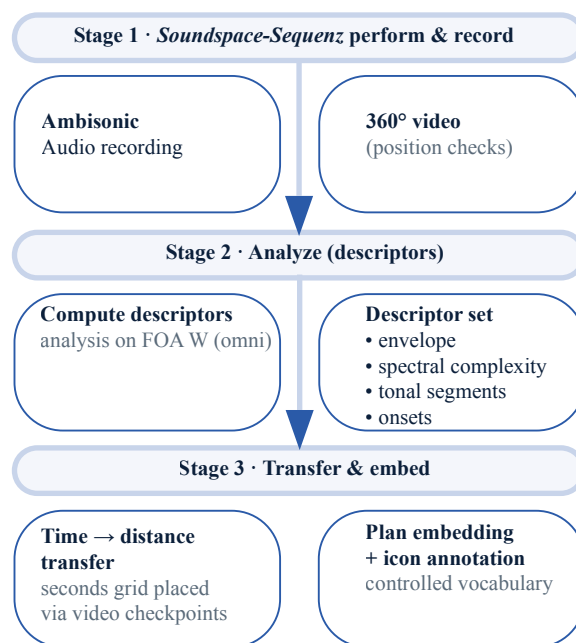


Figure 1. Workflow from traversal to plan-embedded score: FOA + 360° recording, descriptor extraction (FOA W), time-to-distance transfer via video checkpoints, and plan embedding with controlled icon annotation

The workflow has three stages: (1) perform and record, (2) analyze the recorded signal, and (3) transfer time to

distance and embed computed descriptor layers in the plan, complemented by rule-based interpretive icon annotations.

Stage 1: Perform and record (FOA + 360° video)

A *Soundspace-Sequenz* is conducted as an attentive, improvisational walk through the site. Because the practice demands sustained concentration, instrumentation and annotation are kept minimal during the traversal; measurement is deferred to post-processing. The walk is documented using a 360° camera and a first-order Ambisonics (FOA) microphone, enabling subsequent re-listening with spatial and visual context.

Recordings can be replayed either binaurally over headphones or in a multichannel loudspeaker environment. The 360° video provides a positional reference that supports reconstructing turning points, stops, threshold crossings, and other micro-situations, and enables periodic position checks for later alignment.

To ensure comparability across sessions, a basic metadata set is logged (time of day, weather, approximate activity level, and gain settings) and the route is noted. The video functions as a positional scaffold for linking micro-actions (e.g., stopping, turning, crossing thresholds) to audible transitions. This coupling is critical, as the same acoustic event may carry different architectural implications depending on whether it occurs while approaching, passing through, or lingering within a spatial condition.

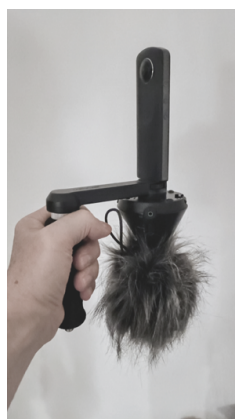


Figure 2. Synchronized recording setup for a *Soundspace-Sequenz*: first-order Ambisonics (FOA) microphone and 360° video camera used for re-listening and positional verification.

Stage 2: Audio analysis (interpretable descriptor set)

For plan-readable descriptor layers within the score, we emphasize descriptors that remain interpretable without specialist signal-processing knowledge and that relate to perceived change: envelope (dynamic contour), spectral complexity (density of spectral peaks), tonal segments (voiced or musical tendency), and onsets (impulsive change points).

In the demonstrated implementation, analysis is run on the FOA W channel (omnidirectional component) to reduce directionality bias in descriptor estimation.

Descriptor settings are tuned for plan readability. We compute features on short analysis windows but aggregate them to a coarser time step (≈ 0.5 – 1 s) so that changes align

with the pace of architectural actions such as walking, stopping, and crossing thresholds. Detection thresholds are set conservatively so the score highlights robust, repeatable features rather than incidental fluctuations of the audio signal.

Stage 3: Fold time into distance, then into plan

The analyzed descriptors are initially time-indexed. To embed them in the plan, time is mapped to walked distance by placing a seconds grid along the path using position checks from the 360° video. This makes accelerations and slowdowns visible through unequal spacing. The scaled path-line is then geometrically aligned with the route and inserted into the plan. In addition to the computed descriptor layers, interpretive icon annotations (atmospheres [10], sound effects [11], soundmarks [12]) are added using a controlled vocabulary.

When embedding the score into a plan, readability is treated as a design constraint. In practice, a contextual plan view is paired with one or more insets where the path-line is shown at notation scale so that descriptor bands and icon annotations can be read without excessive compression. This mirrors musical practice, where an overall form and detailed passages are both necessary for interpretation, and helps maintain the plan's role as a shared medium rather than turning the score into a separate analytic diagram.

IMPLEMENTATION DETAILS

We distinguish between computed descriptor layers (signal-derived and reproducible) and interpretive icon annotations (rule-based and manually assigned), which together form the plan-embedded score.

Signal format and analysis channel choice

A FOA (1st Order) recording contains four channels (X, Y, Z directional components plus W omnidirectional). For descriptor stability across head turns and source directions, the W channel is used as the primary analysis input. The library Essentia was used for the analysis. [13]

Descriptor extraction

Envelope.

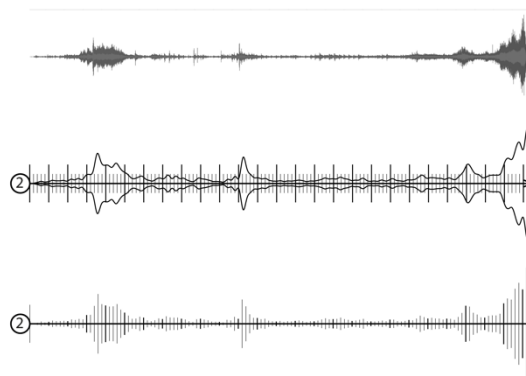


Figure 3. Descriptor example: amplitude envelope (dynamic contour) extracted from FOA W and rendered as a plan-readable layer along the seconds grid.

The envelope represents the loudness trajectory; a seconds grid can be visually weighted by the envelope so that time ticks become salient where the signal is strong.

Spectral complexity.

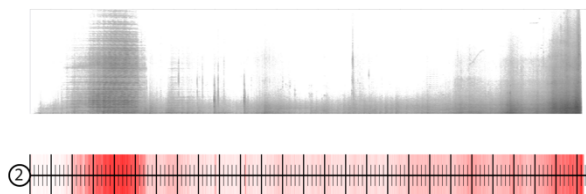


Figure 4. Descriptor example: spectral complexity (density change) rendered as an intensity layer along the seconds grid.

Spectral complexity indicates whether the spectrum is narrow/simple or broad/dense. Results can be drawn as a one-dimensional intensity layer along the path.

Tonal segments.

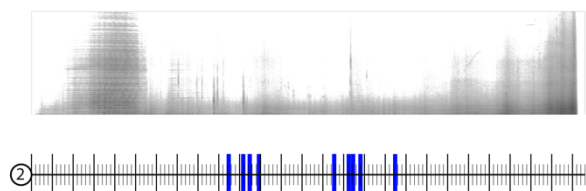


Figure 5. Descriptor example: tonal segments (voiced/tonal material) estimated from spectral flatness and pitch confidence after thresholding.

Tonal segments are estimated from spectral flatness and pitch confidence; thresholding isolates segments likely dominated by voiced speech or musical/tonal material.

Onsets.

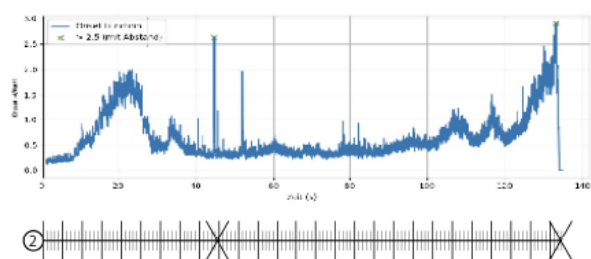


Figure 6. Descriptor example: strong onsets after thresholding, rendered as discrete marks indicating impulsive change points.

Onsets are extracted as peaks indicating sudden change. After thresholding, strong onsets appear as discrete marks along the path, suggesting rhythmic articulation.

Synchronization and temporal alignment

Since the notation links descriptor time stamps to positions in the plan, alignment between audio and video is treated as a dedicated step. The workflow uses the shared recording start as reference and verifies alignment during

re-listening using visible/aural checkpoints such as turns, door crossings, or short impulse sounds. If drift is observed over longer takes, alignment is corrected at a small number of checkpoints so that time-to-distance transfer remains consistent.

Mapping time to path distance

A sequence is simultaneously a duration and a walked length. Mapping time to distance is not assumed linear; instead, the seconds grid is positioned via video observation. This exposes speed changes and supports interpretation: Tick spacing exposes speed variation; onset and tonal layers can then be interpreted against this movement context.

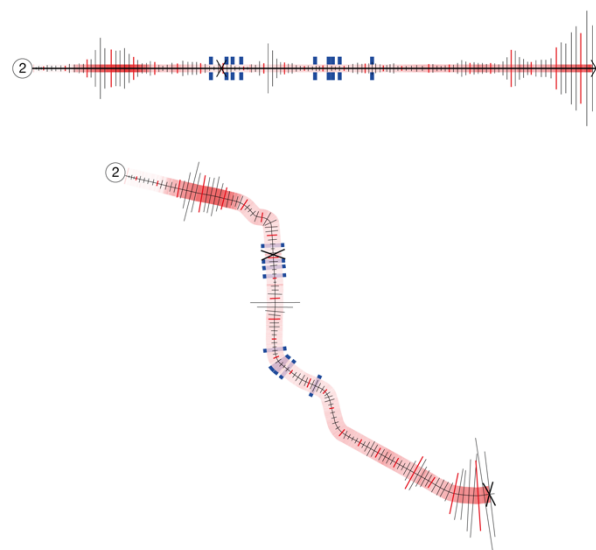


Figure 7. Seconds grid aligned to observed positions (video checkpoints). Tick spacing encodes walking speed; in this example, near-uniform spacing indicates largely steady pace.

Icon layers: atmospheres, effects, soundmarks

Atmospheres are treated as perceivable segments with transitions along the path and annotated using a controlled vocabulary (Figs. 8–10). Each atmosphere segment is labeled with exactly three icons—one for functional setting (e.g., shopping street/park/residential), one for social use (e.g., intimate/communicative/representational), and one for an acoustic descriptor (e.g., open / warm / reverberant / damped)—to maintain legibility and comparability. Sound effects are marked at perceived positions as localized events or short spans using the sound-effect categories listed in [11] (e.g., filtering/occlusion, resonance, reverberant bloom). Soundmarks are placed manually and categorized in line with ISO 12913 categories [1] when they function as identifiable, recurring anchors. Additional effect and soundmark icons may be added beyond the atmosphere label, but only when events remain salient and repeatable during re-listening, keeping the score readable in architectural drawing contexts.



Figure 8. Atmosphere icon set (examples) used to annotate perceptual background segments along the traversal.



Figure 9. Sound-effect icon set (examples) used to annotate localized perceptual transformations (e.g., filtering, resonance, reverberant bloom)



Figure 10. Soundmark icon set (examples) used to annotate identifiable, recurring elements that support orientation.

CASE STUDY: URBAN PUBLIC SPACE TRAVERSAL

Site and recordings

To demonstrate the workflow, multiple *Soundspace-Sequenzen* were recorded in an urban public space. The recording session took place in Spring 2025 in stable weather conditions. Several sequences were captured to show variation across routes and micro-situations.

The site was selected because it contains multiple acoustically distinct micro-zones within short walking distance: a relatively open area, edges exposed to traffic, semi-enclosed corners, and transitional passages where occlusion and reflection change quickly. Such conditions make threshold effects audible and therefore suitable for demonstrating how a scored traversal can capture alternating phases of exposure and shelter that would be flattened by static averaging.

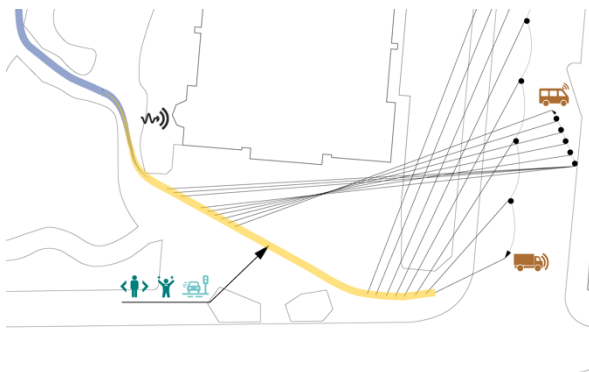


Figure 11. *Soundspace-Sequenz 2 (detail): plan-embedded score with interpretive layers (atmospheres, effects, soundmarks).*

Worked example: Sequence 2

Sequence 2 is used to illustrate the analysis and folding steps. Computed descriptor layers provide the plan-readable analytic backbone of the score: envelope as dynamic contour; spectral complexity as density

transition; tonal segments as voiced or musical moments; and strong onsets as discrete rhythmic articulations. Because the seconds grid is positioned by video-based checks rather than constant-speed assumptions, the drawing can register accelerations, slowdowns, and pauses as changes in tick spacing—information that supports interpretation and comparison across repeated traversals. In the worked example, walking speed was largely steady; accordingly, dense clusters of onsets are read primarily as brief impulse-rich passages rather than as dwell-time effects, while tonal-segment density indicates where voiced/tonal material becomes sustained or recurrent as auditory conditions change along the route (e.g., masking or enclosure), with dense bands suggesting increased presence of voiced sources or acoustically supportive micro-situations and sparse segments indicating predominantly broadband or diffuse backgrounds. The folded representation makes these distinctions visible and supports more precise discussion of where attention was drawn and why.

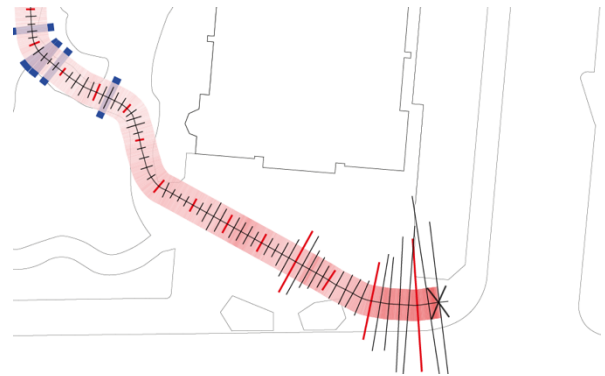


Figure 12. *Soundspace-Sequenz 2 (detail): computed descriptor layers (envelope, spectral complexity, tonal segments, onsets) folded into the plan.*

Plan integration and annotation

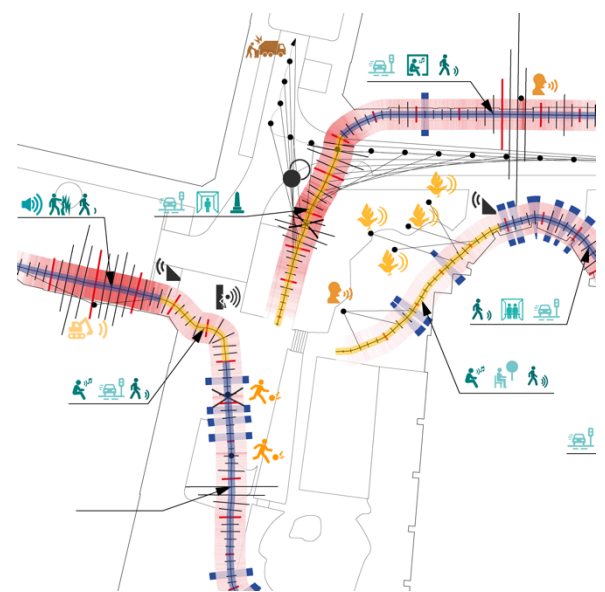


Figure 13. *Soundspace-Sequenzen 1–3 (detail): comparative plan-embedded scores enabling side-by-side reading across traversals.*

After folding, the path is aligned to match its geometry and inserted into the plan. Atmosphere segments are marked, sound effects are plotted, and soundmarks are positioned either on or near the path. The resulting plan-embedded score drawing functions as an audio score in two senses: as a listening guide for re-walking and as an analytic artifact for comparison across sequences.

DISCUSSION

Why this counts as sound as score

Unlike conventional scores that prescribe sound production, the proposed score is derived from situated listening and returned to representation as a guide for re-enactment and discussion. Computed descriptor layers are reproducible from the recordings, while icon annotations capture perceptual categories through rule-based interpretation. The site becomes the instrument, the moving listener the performer, and the notation encodes rhythm and sonic transitions along the path.

The workflow is not intended as a predictive acoustic model or a replacement for compliance metrics; instead, it provides a plan-native representation that supports architectural critique and iteration. Its validation criterion is therefore representational: whether repeated traversals yield scores that remain interpretable, comparable, and useful as guides for re-listening and re-walking within drawing conventions.

Because the score preserves context dependence—movement, orientation, and occupancy—differences between traversals become legible as design-relevant variation rather than being averaged out. At the same time, the method trades spatial exhaustiveness for legibility: a minimal descriptor set and print-oriented smoothing stabilize the score but omit finer-grained directionality and source separation, leaving room for optional extensions when higher analytic resolution is required.

Strengths

- Plan-native communication: inserting the score into plan drawings lowers the barrier to discussing sonic qualities in architectural workflows.
- Interpretability: the descriptor set is intentionally small and legible, supporting comparative reading without black-box metrics.
- Rhythm as design-relevant: the score foregrounds the coupling of movement decisions and sonic transitions as an architectural quality.

In design settings, this plan-native format supports concrete use cases: documenting a baseline listening route during site analysis; comparing repeated traversals across times of day or after an intervention; and supporting review discussions where circulation and material choices can be linked to audible transitions without collapsing the conversation into purely quantitative compliance metrics. Practical use in public space also raises ethical considerations: recordings may capture speech or identifiable persons in the 360° video. In many cases, sharing the plan-based score without distributing raw recordings is sufficient for communication, reducing

exposure while preserving the representational benefits of the method.

Limitations and future work

- Manual annotation load: atmosphere and soundmark annotation remains labor-intensive and depends on expertise.
- Single-channel analysis simplification: using the W channel discards spatial directionality that could enrich the score; future work could add optional spatial layers while keeping readability.
- Scale: the workflow is most effective for micro and meso sites; macro-scale mapping should combine it with broader soundscape methods.

A productive direction is to keep the current layer set as a readable core and add optional modules when needed, such as directional energy measures derived from FOA components or confidence indicators for event localization. This would preserve accessibility for designers while enabling deeper analytic detail in research-oriented applications.

CONCLUSION

We presented a workflow that turns a recorded listening traversal into a path-based audio score embedded in architectural plans. Using FOA + 360° video documentation, a small set of interpretable audio descriptors, and a time-to-distance folding procedure, the method produces drawings that document experience and guide re-enacted listening. The approach extends sound as score from musical production toward architectural perception by scoring a way of hearing a site rather than merely mapping sound.

REFERENCES

- [1] ISO / TS 12913-2:2019(E), *Acoustics - Soundscape - Part 2: Data collection and reporting requirements*, Geneva, 2019.
- [2] B. Leitner, TON : RAUM, Köln: DuMont Buchverlag, 1978.
- [3] M. Neuhaus, *Drawings, Sound Works, Volume 2*, Ostfildern: Cantz Verlag, 1994.
- [4] J. Kang, F. Aletta, T. Oberman, M. Erfanian, M. Kachlicka, M. Lionello and A. Mitchell, "Towards soundscape indices," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen , Germany, 2019.
- [5] M. R. Schafer, Ed., *European Sound Diary*, Burnaby, CANADA: A.R.C: Publications, Vancouver, 1977.
- [6] S. Tanzer, "Soundcities," 1997. [Online]. Available: <https://soundcities.com/info.php>. [Accessed 07 01 2026].
- [7] D. Appleyard, K. Lynch and J. R. Myer, *The view from the road*, Cambridge, Mass.: MIT Pr., 1966.

- [8] G. Debord, "Theory of the Dérive," 1956. [Online]. Available: <https://www.cddc.vt.edu/sionline/si/theory.html>. [Accessed 14 03 2023].
- [9] P. Talwar, *Notation systems in architecture*, Massachusetts Institute of Technology, 1972.
- [10] G. Böhme, *Architektur und Atmosphäre*, München: Wilhelm Fink Verlag, 2006.
- [11] J.-F. Augoyard and H. Torgue, Eds., *Sonic experience: A guide to Everyday Sounds*, Montreal & Kingston, London, Chicago: McGill-Queen's University Press, 2019.
- [12] M. R. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*, Rochester, Vermont: Destiny Books, 1977.
- [13] "Essentia," [Online]. Available: <https://essentia.upf.edu/>. [Accessed 07 01 2026].

SONIC SCORES MIMETIC BEHAVIOUR: TWO CASE STUDIES

Eric Maestri

Sorbonne Université

Institut de recherche en musicologie (IReMus)

eric.maestri@cnr.fr

ABSTRACT

In this paper, I point out questions that have arisen from my practice of sonic scores. This music writing is based on sound. A sonic score is an aural notation that encodes, in a phonographic manner, a musical work in a sound file. By doing so, it places the sonic experience at the heart of musical notation. However, this approach also presents challenges and benefits in musical projects. Sonic scores expand and enhance the musical creativity of performers by freeing them from *graphocentric* approaches. They present musical pieces in a radically different way, offering an alternative perspective on notation and composition. Sonic scores can be used to coordinate musicians and provide a tangible experience with extended instrumental techniques in mixed music. In this paper, I will study the mimetic approach to the performance of sonic scores. I will examine the historical evolution of sonic scores and present an initial analysis of data collected during two sessions of performances of two sonic scores: *Metaction* and *C'est de l'encre*. I will use this experience to make a preliminary evaluation. The goal of this research is to understand the suggested conveyance mode of the proposed sonic scores and to consider them in a broader context.

1. INTRODUCTION

What is a musical relationship? Is interpreting a score comparable to developing a relationship? What exactly is a score? Are scores and notations interchangeable terms? Scores are assemblages of symbols, while notations are methods of writing that link sight and action. Scores are tangible artefacts, while notations are techniques. A score is the arrangement of visual and abstract tools, while instructions for playing are contained within the notation. Scores are published, notations result from centuries of epistemological reflection [1]. Scores are functional to the market, notations are graphical representations. In

Copyright: © 2026 Eric Maestri. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Gregorian chant tradition, notation has a goal: to coordinate a text and a melodic line [2, p. 32]. Those notations are not scores. Scores resulted from an historical and economic process that affected musical production, tastes, and education. The rationalization of music passes through notation [3], and it's nowadays in transformation because of economical and technical development [4]. Scores are a part of this process. They identify the author [5] and give orders thanks to musical education that is industrialized in conservatories. Notation proposes the coordination of visual signs and acoustic facts and signifies musical relationships, pointed out Ligeti [6, p. 154]. However, the listening and performative process cannot be reduced to scores [7]. Notation could be invented and renewed in order to question the standard approach: it is what I try to propose. Now, electronic music modifies the standard approach of musical writing. Fixed media proposes an intermediate format, fixed but not graphical, that allows to create musical works and, meanwhile, demands an open-oral approach to performance that is in between notation and design, considered as the "representation of procedures that allow to produce music and the realization of music through associations." [6, p. 154]

2. THE STANDARD APPROACH

Scores prescribe as an algorithm the construction of a performance [7]. The performer needs to know how to interpret the signs: he must be trained to elaborate a mental act that links visual marks to acoustic occurrences through an instrument [8, p. 50]. The training of the performer is essential for the interpretation of the score. In particular, this kind of training is adapted to scores that indicate the result (*notation of the result*) [6, p. 161]. This last one is a notation that is dedicated to the instruments that are produced by the industry and learned in the conservatories: "the miracle of musical writing [...] depends entirely of the social organization that starts with the industry of the musical instruments, passes through the education and selection of performers [...]" [9, p. 41]. The performer interprets the pitches, the gestures, and the temporality of the piece. This practice defines the tastes and the criteria

of a good performance. Musical writing, differently from articulated languages, is mandatory from the point of view of time. You must read the signs respecting the proportionality and the tempo indicated by the author. The notation used in electronic music indicates the realization of the music from the point of view of its production. It does not indicate the result, the sounding event. Scores must be read and realized in front of a public. Those aspects characterize the standard approach to musical writing, that is *graphocentric*. Composers realize scores as a deferred performance, acts that will be played in the future [10]. Through this approach composers build the “illusion of a possible construction of a language of timbres” [9, p. 29].

2.1 New perspectives

Since 1880, recording techniques and electronic music have opened up new and unexpected compositional spaces. Electronic music allows us to conceive of sound as a composed object, and to record it in the finest detail. Since its beginning, electronic music is an alternative to instrumental music [11]. This evolution responds to the crisis of music in the first part of the 20th century, allowing to control sound as a material [12]. Notions of morphology [13], energy, profile, directionality [14] and sonic images [15] characterize the contemporary compositional vocabulary, determining the new horizon of its control, radically changing the utilization of musical writing. Scores are arrangements of sounds and constructs of sonic sculptures. Each note is an “atom” of music [16]. In this context, notes are used to control much larger and more complex sonic elements. They are signs in the context of software; they coordinate an encoded architecture: they control what you want, such as samples, synthesis, neural networks, and spectra. You can decipher a score or use it to create your desired output. Producing a sound with a computer means to control the vibrations of a loudspeaker membrane [17]. Musical writing means must be performative [18, p. 3]. Writing technologies can represent what is perceived and generate sonic events from those signs.

3. SONIC SCORES

Sonic scores use aural notation and encoding, in a phonographic manner, the musical work in a sound file. This approach is a recent tradition, starting in the 1990, but has been present in electronic music practices since the 1970s. In 1978 Denis Dufour transcribed Pierre Schaeffer’s *Étude aux sons animés* for the synthesizer group TrioGRM+. The piece *Souvenir de Pierre* was performed live and is based on the fixed-media piece realized by Pierre Schaeffer in 1958. Schaeffer has the same idea since 1948. He planned to write a new work starting from *Étude noire* and *Étude pathétique* (1948) [19, p. 35]. These pieces could be used as models for instrumental writing. The porosity of instrumental and

electronic practice defines the relationship of the musical tradition with new technologies. Alain Savouret conceived the “loudspeaker pieces” (*pieces haut-parlantes*). These pieces have a pedagogical objective. They confront the musician interpreter with “very different sonic energy sources” [20]. Sandeep Bhagwati invests this field conceiving *audio scores*. This term “denotes a type of score that uses headphones as its interface to the musician and conveys musical information primarily via acoustical messages.” [21, p. 25]

I have created my own method in this field, harnessing the expressive power of electronic music. I challenge the performer to a critical threshold, prompting them to invent novel instrumental sounds and techniques. Electronic music has the ability to shift perspectives and liberate artists’ creativity. Performers must play what they hear, not what they read. The sonic score is an invisible score with sound affordances, ‘morphocons’ [22]. I realized sonic scores with electronic sounds in order to engage the limit of mimesis [23]. This “conveyance mode” asks the performer to imitate a “sound heard in the headphone” [21, p. 28]. This choice creates a fascinating divide between the sonic symbol and the musical outcome, a divide that cannot be bridged, leading to a powerful tension in the performance. This approach reminds the results of the notation in “new complexity” [24]. This musical concept is presented in a way that is both meticulous and enigmatic, inviting infinite interpretations. I prescribe a sound through a sound; the performer must play the sound that I would like to hear, acknowledging that it is utopian. This writing technique enhances the artist’s creativity, leading to an unanticipated bond between the composer and the performer [25].

Compositional criteria underpin this approach. I created the sound scores by seeking to activate the link between sound and gesture. I am thus part of this line of thinking that cuts across musicology and musical practice in the 20th and 21st centuries. Denis Smalley designs gestural archetypes that form the basis of his spectromorphological theory [26]. Denis Smalley emphasises the relationship between bodily gestures and sounds, which are initially involuntary, then instrumental and electronic. Sounds are increasingly distant substitutes for everyday bodily movements and those of the environment. This hypothesis links sound, movement and context, allowing us to think about the score within this chain of concepts. It is also based on cognitive science: according to Schleidt and Kien [27], the average duration of human actions is three seconds, which, for Godøy, corresponds to the duration of the ‘balanced’ sound objects indicated by Pierre Schaeffer [28, p. 57]. These objects determine a temporal dimension that has the same duration as human movements, different from that expressed by “eccentric sounds” [29, p. 437] which, on the other hand, exceed this deep structure linking sound to the body. Based on this interpretation of Smalley’s spectromorphology, I prepared electroacoustic

sequences that could be interpreted by instrumental gestures as scores.

4. TWO CASE STUDIES

Since 2019, I have applied this approach in various ways. The most ambitious project to date is *Ora* (2024), a commission and performance by the ensemble HanatusMiroir. I specifically designed the sonic score for flute, violin and percussion, with the percussion part written to achieve a range of timbres and materials. In this paper I will discuss *Metaction* composed for flute, harp and viola in 2020 and *C'est de l'encre*, written for viola and piano in 2023, performed by Grazia Giacco and Raffaella Valente [25].

4.1. Methodology

The analysis of these excerpts is based on the recording that the musicians and I made. I looked for the connection between electronics and the instrument in both cases. In the first case, the performers had to play the score, imitating its sounds: just as notes, they were signs. I tried to understand the type of relationship between the electronic part (the sonic score) and the instrument. I propose that we analyse mixed sounds as a conservative emergent property by using the concept of functional partitioning [30, p. 57]. This concept describes how the organization of instrumental and electronic sounds works harmonically and temporally. The elements are divided and have a function in the resulting sonic whole. The temporal functional division is the horizontal arrangement of both electronic and instrument components. The vertical axis also has a spectral functional division. This methodology allows the understanding of the functional partitioning of both sound dimensions and their role in the sound. A similar remark characterised an approach proposed by Laura Zattra as well [31]. This relationship can be interpreted as *isomorphic* or *heteromorphic* [30]. Recording sessions were held with participants. *Metaction* was performed thirteen times, while *C'est de l'encre* was performed four times. The first one is played by soloists while the second by a group of ten musicians.

4.2. Metaction

Organized by Ensemble l'Itinéraire, OpusOne is a music academy for performers interested in trying their hand at composition. They can test their first sketches and pieces with other musicians. The participants are all professional performers looking to incorporate musical writing into their practice. For this reason, most of them are improvisers. I had the opportunity to serve as a tutor in the 2025 edition of the Musical Academy. One of the tasks I assigned to the students was to interpret a sonic score, specifically *Metaction*. The score consists of three parts: the first for plucked or bowed strings, the second for

bowed strings, and the third for winds and brass. The musical composition consists of two distinct parts:

1. Sudden electroacoustic gestures and environmental noises (birds in particular);
2. A progressive and slow glissando towards the lower register;

The score is designed to elicit different extended instrumental techniques through various sound configurations. Originally written for flute, viola and harp, it is open to interpretation by any musician. The musicians were given two weeks to submit their interpretations of the score.

GD: Singer. The performer uses the score to create something unique, playing each part and then mixing them. The performer realized a very detailed interpretation, preserving gestures and pitches proposed by the sonic score. [*Isomorphic*]

FF: Cellist and singer. The performer realized the score with the cello and the voice. The modality of the interaction emphasized the dynamic contrasts and the figures of the piece. The performance was expressive and the cello and voice created a unified texture. [*Isomorphic*]

AF: Saxophonist. The performer played the score precisely, capturing all the sonic details that the score proposes. [*Isomorphic*]

EL: Pianist. The performer chooses the part conceived for plucked instruments and plays the score, alternating between interpreting it with the keys and the piano's soundboard. [*Heteromorphic*]

FM: Tubist. The performer creates a unique version using the contrasting registers of the instrument. The performer follows the score, forcing the instrument and its techniques to play the prescribed sound. [*Isomorphic*]

GM: Trombonist. The performer uses, differently from the others, electronic treatments and amplification. The player performs the score as prescribed. [*Isomorphic*]

AN: A pianist. As EL, the artist attempts to investigate the soundboard and the keys, combining these two methods. [*Heteromorphic*]

DN: Pianist. The performer delves into the expressive potential of the piano, deviating from the notated gestures. The resulting sounds are distinctly different from the intended sonic output. [Heteromorphic]
ON: Pianist. The performer plays a study that focuses on the soundboard, exploring its resonance and the density of the movements. The performer uses extended piano techniques, such as clusters on the keys. The piece is intense and aggressive, ending with whistles. [Heteromorphic]
MP: Trumpet player. The performer expands the techniques of the instrument from the lowest to the highest register, mutating the traditional timbre of the instrument. [Isomorphic]
DT: Flutist. The performer realizes a poetic version of the piece, independent of the result. The result is partially detached from the prescribed sound but respects the form and the global dynamic. [Heteromorphic]
TP: Guitar. The performer uses a guitar and plays slight glissandos on the strings, rough notes, pointillistic tones. [Isomorphic]
XS: Violinist. The performer plays following the score, alternating the techniques and gestures. [Isomorphic]

Musicians playing the piano adhere to the gestures, creating a metaphorical version of the score. Due to the concrete limits of the instrument, the outcome is heteromorphic compared to the intended sound. The other musicians, taking advantage of their instrument's versatility, perform the sonic instructions as directed. The proposed composition alternates between loud and quiet sections, seamlessly blending movements and sounds with the original score.

4.3. *C'est de l'encre*

The piece, initially written for viola and piano, is used in the context of a group free-improvisation. In the winter of 2025, I recorded four impromptu sessions with a group of ten musicians as part of a lesson on arranging a new composition. This approach is inspired by improvised music with tape [32]. To introduce the group to the world of extended techniques, while their knowledge of contemporary music was still nascent, the use of sonic scores was employed to facilitate improvisation and the discovery of new techniques. The composition, lasting six minutes, consists of four distinct parts, each alternating between gestural and textural segments. The first section,

marked by a gradual and directional movement, is characterized as being gestural.

1. The second part consists of long, sustained, synthesized sounds.
2. The third part blends the first two, based on a steady rhythm: movements and textures.
3. The final part is melodic, with a repeated motif in the upper register.

The group's instrumentation includes a flute, trumpet, saxophone, two electric guitarists, an electric bass, a cello, and two pianists.

Performance I. In this performance the group starts listening to the piece for the first time. They try to coordinate themselves. As they uncover the musical notation, they realize that they must perform by mimicking this reference point. The sonic score encourages the musician to ad-lib in an environment that provides tonalities, outlines, and motions. In response, the performers interact with one another, creating a musical feedback loop.

Performance II. Players know the score already and start to play with the recording and the new musical configurations. The plucked-string instruments provide a contrasting counterpoint to the first section with their short, staccato events. The second part is characterized by repetition, long notes and pitch variations. The group follows the crescendo of the tape. The dynamic progression has an important impact on improvisation. In the final part, the musicians follow the melody of the tape.

Performance III. In this performance the musicians improvise, creating noises; the performance is more complex. Players know the score and anticipate its form. They align with the music of each part, counterpointing the tape with noisy gestures in the first segment and following the textural evolution of the music from a register and content perspective in the second. The third rhythmic part is performed with counterpoints and repetitive gestures. This part finishes with a climax of the group making noises and verses with their voices. After a pause, the final section consists of fragmented sound events following the melodic profile proposed by the electronics. In this part, the musicians engage in instrumental dialogue and counterpoint with the electronics.

Performance IV. The last improvisation is characterized by counterpoints and gestures, especially in the first section. Now that the musicians know the score, they

play with it, fully respecting the form. Musicians realize inside jokes and free their playing. In the second part, they interpreted the music based on texture. The third part is characterized by melodic counterpoints that prepare for the final part. In this performance, the final part is melodic and high-pitched, and it is accompanied by dense harmonies and gestures in the lower register, giving a new version of this final part.

5. CONCLUSIONS

The performances I reviewed shed light on the practice of sonic scores. In Metaction, the sonic score encourages the performer to modify their instrumental technique. The main “sonic behaviour” [21, p. 29] is characterized by mimesis. Performers play the sonic score, trying to realize something utopic on their instrument. The instrument used by the musician changes the approach. Instruments that can produce microtonal modulations and alterations in sound over time have the ability to mimic. For the piano, the relationship is metaphorical and requires more invention. The sonic score expands the possible range of instrumental notation. In *C’est de l’encre*, the performers play in groups improvising. As the experience progresses, the group becomes aware of the score. If, at the beginning, they try to understand the form and the sounds, at the end they know it and play the score playing between them. The result is a composed improvisation based on the trace of the electronics. Electronics create an environment for interaction, forcing the performer to find solutions. Sonic scores propose an alternative to graphocentric approaches, obliging the performer to coordinate their playing with an often unrealizable sound. This expands their creative potential. Sonic scores can be viewed as a component of the creative process for musicians, providing them with the autonomy to generate music within the framework of a predetermined musical composition. This expands the boundaries of the composer’s function, facilitating the creation of open musical pieces and innovative notation that incorporate the multimedia aspects of musical writing.

Acknowledgments

I would like to thank the participants in this experience: the composers and performers from OpusOne 2025, organized by Ensemble L’Itinéraire, and the postgraduate students in music education at Sorbonne University, Paris. This research project was initially a personal endeavour, but the involvement of these talented musicians has allowed it to expand and evolve.

6. REFERENCES

- [1] M.-E. Duchez, ‘An historical and epistemological approach to the musical notion of “form-bearing” element’, *Contemporary Music Review*, vol. 4, no. 1, pp. 199–212, 1989.
- [2] S. Rankin, ‘Écrire les sons. Création des premières notations musicales’, *Cahiers de civilisation médiévale [En ligne]*, vol. 2020, pp. 379–392, 2012, doi: <https://doi.org/10.4000/ccm.10994>.
- [3] M. Weber, *Sociologie de la musique. Les fondements rationnels et sociaux de la musique*. Paris: Metaillé, 1998.
- [4] H. Lehmann, *La révolution digitale dans la musique. Une philosophie de la musique*. Paris: Editions Allia, 2017.
- [5] N. Goodman, *Languages of Art. An Approach to a Theory of Symbols*. Indianapolis/Cambridge: Hackett Publishing Company, Inc., 1976.
- [6] G. Ligeti, ‘Nouvelle notation. Moyen de communication ou but en soi?’, in *Neuf essais sur la musique*, Genève: Contrechamps, 2010, pp. 153–168.
- [7] C. Small, *Musicking. The Meaning of Performing and Listening*. Middletown, Connecticut: Wesleyan University Press, 1998.
- [8] P. Hernadi, ‘Reconceiving Notation and Performance’, *The Journal of Aesthetic Education*, vol. 25, no. 1, pp. 47–56, 1991.
- [9] J.-M. Chouvel, *Partitions*. Sampzon: Delatour France, 2025.
- [10] E. Maestri, ‘Notation as Temporal Instrument’, in *Proceedings TENOR 2016. International conference on technologies for music notation and representation*, R. Hoadley, C. Nash, and D. Fober, Eds, Cambridge, May 2016, pp. 226–228.
- [11] K. Stockhausen, ‘Elektronische und instrumentale Musik’, in *Texte zur elektronischen und instrumentalen Musik, Aufsätze 1952–1962 zur Theorie des Komponierens*, vol. 1, Köln: Dumont Schauberg, 1963, pp. 140–151.
- [12] I. Xenakis, ‘La crise de la musique sérielle’, *Gravesaner Blätter*, vol. 1, pp. 2–4, 1955.
- [13] K. Stockhausen, ‘The Origins of Electronic Music’, *The Musical Times*, vol. 112, no. 1541, pp. 649–650, 1971.
- [14] A. Vande Gorne, ‘Space, Sound and Acousmatic Music, the Heart of the Research’, in *Compositions for audible space*, M. Brech and R. Paland, Eds, Berlin: Music and Sound Culture, 2015, pp. 205–220.
- [15] D. Smalley, ‘The Listening Imagination: Listening in the electroacoustic era’, *Contemporary Music Review*, vol. 13:02, pp. 77–107, 1996.
- [16] T. Murail, ‘The Revolution of Complex Sounds’, *Contemporary Music Review*, vol. 24, no. 2/3, pp. 121–135, 2005.
- [17] M. Mathews, ‘The Digital Computer as a Musical Instrument’, *Science*, no. 3591, pp. 553–557, 1963.
- [18] A. Veitl, ‘Notation écrite et musique contemporaine : quelles grandes caractéristiques des technologies numériques d’écriture musicale’, in *Actes des Journées d’Informatique Musicale 2007*, 14 avril 2007. [Online]. Available:

<http://www.tscimuse.org/biblios/veitl/technologiesecrituremusicale.pdf>

- [19] P. Schaeffer, *À la recherche d'une musique concrète*. Paris: Éditions du Seuil, 1952.
- [20] A. Savouret, *Introduction à un solfège de l'audible*. Lyon: Symétrie, 2010.
- [21] S. Bhagwati, 'Elaborate Audio Scores: Concepts, Affordances and Tools', in *Proceedings of the 4th International Conference on Technologies for Music Notation and Representation*, S. Bhagwati and J. Bresson, Eds, Montréal: Concordia University, 2018, pp. 24–32.
- [22] G. Parsehian and B. F. Katz, 'Morphocons: A new sonification concept based on morphological earcons', *Journal of the Audio Engineering Society*, vol. 1, no. 1, pp. 1–10, 2011.
- [23] S. Emmerson, *The language of electroacoustic music*. Basingstone: Palgrave Macmillan, 1986.
- [24] B. Ferneyhough, 'Il tempo della figura', in *Collected Writings*, J. Boros and R. Toop, Eds, Amsterdam: Harwood Academic Publishers, 1995, pp. 33–41.
- [25] E. Maestri, G. Giacco, and R. Valente, 'Les partitions sonores. Un outil collaboratif entre l'interprétation et la composition', *Revue musicale OICRM*, vol. 12, no. 1, 2025.
- [26] D. Smalley, 'Spectromorphology: explaining sound-shapes', *Organised Sound*, vol. 2, no. 2, pp. 107–26, 1997.
- [27] M. Schleidt and J. Kien, 'Segmentation in behavior and what it can tell us about brain function', *Human Nature*, vol. 8, no. 1, pp. 77–111, 1997.
- [28] R. I. Godøy, 'Gestural Affordances of Musical Sound', in *Musical Gestures. Sound, Mouvement, and Meaning*, New York: Routledge, New York, 2010, pp. 103–125.
- [29] P. Schaeffer, *Traité des objets musicaux. Essais interdisciplines*. Paris: Editions du Seuil, 1966.
- [30] E. Maestri, *Contacts. Essai sur la musique mixte*. Louvain-la-Neuve: EME Éditions, 2023.
- [31] L. Zattra, 'Les écritures des musiques informatiques mixtes à la lueur des sources documentaires', in *Regards sur les musiques mixtes*, Marc Battier., Paris: INA-GRM, 2016, pp. 195–230.
- [32] P. A. Tremblay, 'Mixing the Immiscible: Improvisation within Fixed-Media Composition', in *EMS 2012 Proceedings*, Stockholm, 2012.

SONGS OF LOVE AND DEATH: THE ACT OF SOUND AS NOTATION

Gil Dori

ENTI-UB

`gil.dori@enti.cat`

ABSTRACT

Songs of Love and Death is a composition for voice and electronics, which revolves around vowel sounds - real and implied. It features a real-time graphic score, which uses mouth contour data to recreate the shapes of vowels on the screen. Although they are represented graphically, these recreated vowels imply the sound that is tied to each one of them. The performer sings the text, imitating the given vowel shapes, and thus reshaping the sounds of the sung text. Through that, the work explores unusual acoustic results that emerge from the contrast between the implied vowel sound and the text needed to be sung. This approach to action-based notation incorporates aspects of audio scores into a visual one. The paper describes the composition and performance of *Songs of Love and Death*, discussing this notion of implied sounds, or acts of sound, as the driving principle of the notation.

1. INTRODUCTION

1.1. Action-Based Visual and Audio Scores

One of the types of real-time scores concerns notating gestures and action. The underlying notion of such action-based notation is to enable quick, accurate interpretation on the spot, through clear, immediate instructions [1]. Often action based scores are notated graphically, using visual symbols to represent gestures, but some utilize audio messages to convey performance instructions and information of musical gestures [2, p. 2].

Visual action-based scores rely on sound morphology. That is the gestural structure of sound manifested in time, an inherited physical interrelation between visible gesture and sound [3]. This allows the performers to instantaneously and intuitively interpret the score by imitating the given action. When coupled with a gesture data approach to notation, it produces results that are completely idiomatic to the instrument [4]. Imitation is also an important interpretation method of audio scores. However, as opposed to the visual score, audio score

imitation embraces performance possibilities that fall outside the idiomatic practice, producing results that would not be habitually performed [2, pp. 4-5].

The composition discussed in this paper, *Songs of Love and Death* (2025), uses an action-based visual graphic notation. It features actions on the screen, recreated in real-time directly from gesture data, which the performer needs to imitate. It also incorporates the audio score notion of eschewing idiomatic performance practice. It aims to be intuitive as well as to produce unconventional performance and acoustic results. While not featuring actual audio instructions, it utilizes implied sounds, or the act of sound, to achieve this outcome. This is done in the context of vowels, which lend themselves naturally to the idea of implied sounds.

1.2. Audio-Visual Perception of Vowels

Vowels are speech sounds made with an open vocal tract, without significant restriction of the air flow, and with specific tongue and lips position that define the sound through filtering. This can be measured through spectral analysis, showing the formats that make the vowel sound the way it is [5]. Vowels, then, are an act of sound that presents a distinct morphology. Only seeing the act of producing vowels implies the sound of the vowel. As opposed to the understanding of the morphology of an instrument, which requires specific training, the morphology of vowels is understood naturally and intuitively.

McGurk and MacDonald, in their seminal paper *Hearing lips and seeing voices*, illustrated our innate ability to interpret vowels by both audio and visual information [6]. Other researchers have further studied this phenomenon, dubbed the McGurk Effect. While results vary, depending on the vowels and languages, the audio-visual perception of vowels is clear [7]. Conrad even determined that the difference in lip reading between deaf and hearing 15-year-olds is negligible [8], and Havy et al. demonstrated that not only both toddlers and adults can recognize words visually, but adults can also even learn new words this way [9].

Songs of Love and Death, in a way, is a live musical performance of the McGurk Effect. Its notation shows the mouth shapes and the text, each with its implied vowel

sound. But the acoustic result is a fusion of both. Or, in audio terms, considering the vowels as filtered sound, it is akin to convolution, in which the characteristics of an implied sound is used to alter the produced sound.

2. COMPOSITION PROCESS

2.1. Key Elements

Songs of Love and Death is a composition for voice and electronics, with a real-time graphic score. The score was made in Processing. Some of the audio material was made in Pure Data, and gesture data acquisition was done with FaceOSC.

The three elements—the signer, the synthesized accompaniment, and the notation—are all united through vowels. The composition is defined by the relationship between the act of producing vowels and the acoustic results of the actual produced vowels. It explores the contrast between the implied vowel sound and the text needed to be sung. The vowels used are the five most basic ones: open front unrounded ⟨a⟩, close-mid front unrounded ⟨e⟩, close front unrounded ⟨i⟩, close-mid back rounded ⟨o⟩, and close back rounded ⟨u⟩ [10].

The singer, could be any type of range, is instructed to sing a given text from various poems about love and death. The text is fragmented, and presented as isolated short phrases, words, and individual syllables. The score shows the text one line at a time. The text is fixed, and is introduced in the same order, so it can be learned in advance. The graphic notation indicates relative pitch, dynamic, and most importantly, mouth shape to imitate. When singing the text, the singer must produce its vowels according to the given mouth shape (figure 1).

The electronics accompany the singer with synthesized vowels and fricatives, utilizing formant synthesis and source-filter modeling respectively. Initially I tried to have all the audio synthesis happen in real-time and have it done in Processing, but in the case of the vowels it yielded poor results, both aesthetically and functionally. Instead, I synthesized the vowels in Pure Data, with added vibrato and reverb, recording four sound files for each of the five vowels in variation of pitch and duration. The audio files, then, were called up by the score algorithm when needed, and, in some cases, played at different rates to alter the pitch. In the composition, the synthesized vowels function as sustained choir-like pads, with long attack and release. In contrast, the synthesized fricatives are more abrupt, and in some cases, they are played all together randomly, as a rapidly changing texture. They are synthesized in Processing in real-time. Four types of fricatives are synthesized: voiceless labiodental ⟨f⟩, voiceless dental ⟨θ⟩, voiceless alveolar ⟨s⟩, and voiceless postalveolar ⟨ʃ⟩ [Ibid.]. Overall, the electronics accompaniment is sparse, and it is used mostly to emphasize certain moments in the piece.

The most prominent component of the score is the vowel shape indication. This is the act of sound which informs the voice production of the performer. Essentially, an animated mouth that changes according to the vowel selected by the score algorithm. I used

FaceOSC to record mouth shape data of each of the five vowels into text files. Then, in Processing, I recreated a mouth with the same points as in FaceOSC, and that is capable of the same movements. For this process I made use of the FaceOSC Receiver Templates repository by Wilcox et al. [11]. Finally, when a vowel is selected by the program, it reads the data from the corresponding text file, and the mouth changes its shape accordingly (figure 2).

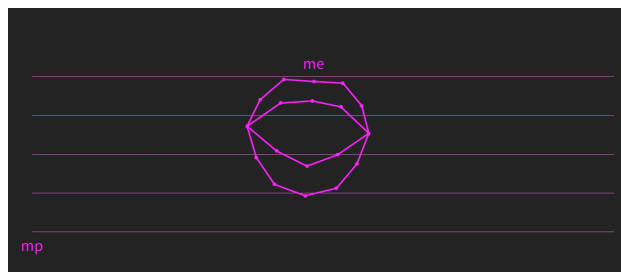


Figure 1. *Songs of Love and Death* score, showing the mouth shape indication, text to be sung, pitch and dynamic indication.

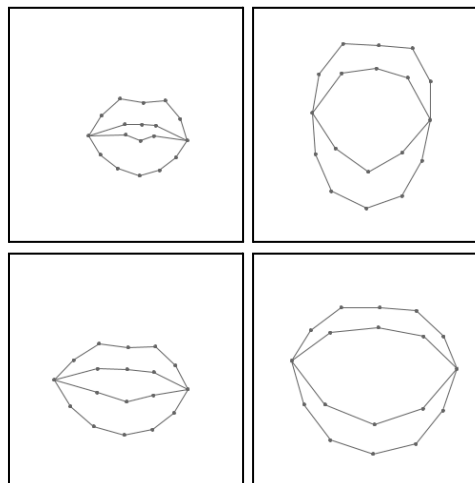


Figure 2. Preliminary mouth shapes recreated in Processing from FaceOSC data.

2.2. Score Design

In the score itself, the animated mouth also moves vertically, to indicate the pitch. A higher position corresponds to a higher pitch, and vice versa. While pitch is indeterminate, the mouth is placed over a five-line staff (without a clef), and it moves within the common singable range from slightly below the bottom line to slightly above the top line (figure 3). Dynamics are indicated by the color of the background; the brighter it is, the louder the level, and the darker it is, the softer the level. Dynamics are also indicated by traditional letter markings, below the staff on the left (figure 4). Initially, the score did not include the staff and dynamic markings, but they were added during the testing process to facilitate performing the piece. The purple color of the whole score was also proved to be the most visually clear,

taking into consideration the changing brightness of the background.

The text is placed directly above the mouth, which helps with visual focus. The performer is instructed to sing only when text appears. Rests are indicated by a closed mouth in the center of the score. Just like lines of text, the placement of rests is also fixed. The duration of each line and rest, though, may vary. There are three fixed long rests that mark the endings of sections. Overall, there are four sections to the piece, defined by text grouping. With each section, the text becomes more fragmented.

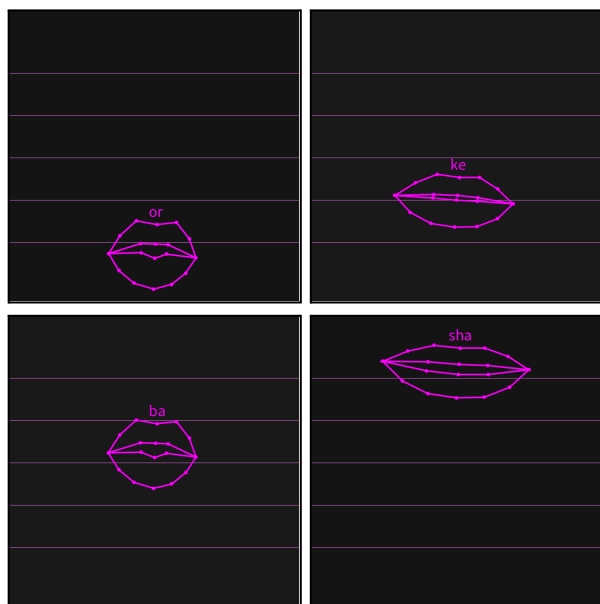


Figure 3. Pitch indication and range, from lowest (top left) to highest (bottom right).

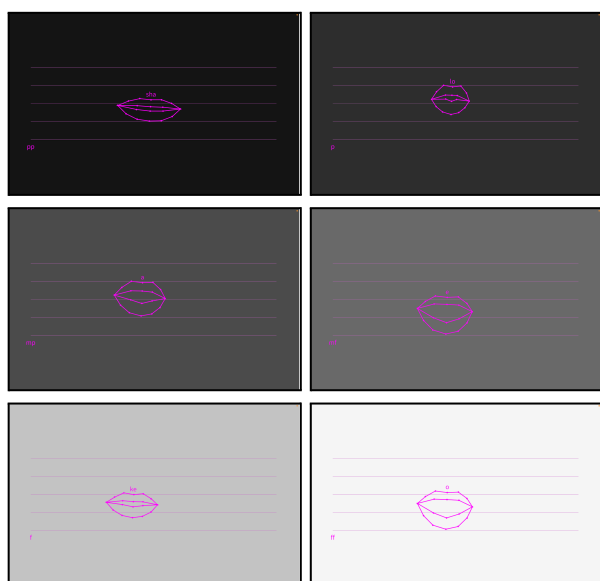


Figure 4. Dynamics indication and range, from *pp* (top left) to *ff* (bottom right).

The score is controlled automatically by an algorithm that selects which vowel shape to display, as well as which synthesized vowel to play, including in what pitch. The selection is done similarly, using first-order Markov chains, but according to different inputs. In the case of visual display and audio playback, vowels are selected according to the text. The selection of pitch is done according to pitch tracking of the performer's input. In addition, an envelope follower is used for smoother control over the electronics' amplitude. However, in various places the automatic selection is overridden in favor of fixed compositional decisions, to better shape the piece.

2.3. Songs of Love and Death in Practice

Bringing the piece onto the stage was done in two phases. The first was a preliminary test performance that took place in a small venue, performed by a soprano. It allowed better preparation for the second phase, which included three performances in concerts of a more traditional, official context by a baritone singer. A video recording of the performance is available online [12].

The graphic score itself proved intuitive and easy to follow, especially in the second phase, after the rounds of testing and improvements during the first phase. In the second performance, the singer also printed the text, to learn and anticipate it better, which further improved the performance.

In both phases, the main challenge was working with the singers on how to approach vocal production. Eventually, the approach that worked the best was to extremely exaggerate the mouth shape given by the score when singing the text, and to consciously prioritize the acoustical result over the text, even up to not attempting to make the text intelligible at all. While both singers naturally understood the morphological aspect of the implied sound instructions, that is the act of sound given by the animated mouth on the screen, applying it into their performance practice required a lot of effort. In this way, the score functions closely to an audio score, rather than a visual one. Not only that the performers needed to embrace possibilities and outcomes that are foreign to their idiomatic vocal practice, but they also had to completely abandon the way they usually present text in song.

3. CONCLUSIONS

Songs of Love and Death was well received by the public, and most importantly, by the performers. Both singers concluded that performing the piece was a valuable experience worth the effort.

The composition successfully shows that an act of sound, or implied sound, approach can be beneficial within the context of visual action-based score. On the one hand, it is intuitive to understand and follow. On the other hand, it embodies audio score notions of achieving unfamiliar acoustic outcomes.

This approach to notation may prove relevant to real-time scores in general, inviting further artistic and research work using similar procedures. Future development possibilities may include having an additional performer performing the act of sound live, directly controlling the score; adding an artificial intelligence vowel detection tool, for more natural algorithmic decision making; and conducting research into audience perception of this type of notation.

4. REFERENCES

1. S. Shafer, "Performer Action Modeling in Real-Time Notation," in *Proc. Int. Conf. on Technologies for Music Notation and Representation – TENOR'17*, Universidade da Coruña, 2017, pp. 117-123.
2. S. Bhagwati, "Elaborate audio scores: Concepts, affordances and tools," in *Proc. Int. Conf. on Technologies for Music Notation and Representation – TENOR'18*, Concordia University, 2018, pp. 24-32.
3. T. Wishart, *On Sonic Art*. Routledge, 1996.
4. G. Dori, "Using gesture data to generate real-time graphic notation: A case study," in *Proc. Int. Conf. on Technologies for Music Notation and Representation – TENOR'20*, Hamburg University for Music and Theater, 2020, pp. 68-74.
5. S. A. Fulop, *Speech Spectrum Analysis*. Springer Science & Business Media, 2011.
6. H. McGurk, and J. MacDonald, "Hearing lips and seeing voices," *Nature*, 264, pp. 746–748, 1976, doi: <https://doi.org/10.1038/264746a0>
7. S. Shigeno, "Influence of vowel context on the audio-visual speech perception of voiced stop consonants," *Japanese Psychological Research*, vol. 42, no. 3, pp. 155–167, 2000, doi: <https://doi.org/10.1111/1468-5884.00141>
8. R. Conrad, "Lip-reading by deaf and hearing children," *British Journal of Educational Psychology*, vol. 47, no. 1, pp. 60–65, 1977, doi: <https://doi.org/10.1111/j.2044-8279.1977.tb03001.x>
9. M. Havy, A. Foroud, L. Fais, and J.F. Werker, "The Role of Auditory and Visual Speech in Word Learning at 18 Months and in Adulthood," *Child Development*, vol. 88, no. 6, pp. 2043-2059, 2017, doi: <https://doi.org/10.1111/cdev.12715>
10. International Phonetic Association, *Interactive IPA Chart in English*. Accessed on: January 15, 2026. [Online]. Available: https://www.internationalphoneticassociation.org/IPAcharts/IPA_charts_TI/IPA_charts_TI.html#eng
11. D. Wilcox, K. McDonald, G. Levin, A. Carabott, D. Moore, and S. Gruber, *FaceOSC Receiver Templates*. Accessed on: January 17, 2026. [Online]. Available: <https://github.com/CreativeInquiry/FaceOSC-Templates/tree/master>
12. G. Dori, *Songs of Love and Death*, 2025. Accessed on March 15, 2026. [Online]. Available: <https://youtu.be/dbAC6nXrnVI>

INTERACTIVE SCORES: TASKS AND DECISIONS IN IMPROVISATION SYSTEMS AND OPEN WORKS

Per Anders Nilsson

Academy of Music and Drama
University of Gothenburg
per.anders.nilsson@hsm.g
u.se

Palle Dahlstedt

Academy of Music and Drama, and
Dept. of Computer Science & Engineering,
University of Gothenburg and Chalmers
palle.dahlstedt@gu.se

ABSTRACT

This study investigates relationships and interactions between a performer and a score. This view implies a shift in focus, from the score to the act of interpreting and performing, and subsequently to the performer. A significant aesthetic conception is to regard the score as an artifact intended for players to interact with, rather than a means for personal expression per se. The musical context is first and foremost within the experimental music tradition in the post-World War II era. The theoretical framework consists of concepts from, e.g., John Cage and Cornelis Cardew, in addition to influences from the field of interaction design. To elucidate the subject matter, a limited number of musical works are discussed, including both classic pieces and original works. Finally, the paper concludes with some personal reflections on working with interactive scores.

1. INTRODUCTION

The aim of this study is to investigate relations and interactions between a performer and a score, before and during a performance, including explicit and implicit interpretative freedom. The aesthetic and theoretical context is primarily within the experimental music tradition in the post-World War II era, including the concept of an open work. This view implies a shift in focus, from the score to the act of interpreting and performing, and subsequently to the performer. One may approach scores in many ways, e.g., in relation to musical outcomes, as carriers of musical ideas over time and space, as simply a set of instructions, or as autonomous artifacts. Furthermore, in our view, it is not possible to discuss scores without also addressing the concept of a work; these

concepts are entangled. Yet another issue that will be elucidated is the question of how little information a score might contain to count as a score. That is, to give participant players necessary information of how to interpret and perform a musical piece, that may be repeated, with recognizable and preserved identity. It is worth noting, however, that there are many traditionally oriented composers and musicians of today who are not interested and/or comfortable with the mandatory freedom in open works, and therefore are not the object of this text.

Concepts and Notions

A significant aesthetic conception in this study is to regard the score as an artifact intended for players to interact with, rather than as a means for personal expression per se. Likewise, our personal experience as practitioners, composers and musicians in this tradition plays a major role. In this context, a composer is regarded as an interaction designer. Swedish interaction researchers Lars Hallnäs and Johan Redström claim that: “Interaction design is the design of the interaction between users and products”¹. Accordingly, the interaction design perspective shifts the focus from the producer to the user, and in this context from the composer to the performer:

So one central leitmotif here is that interaction design introduces a shift of focus from the things themselves to the acts that define them in use. This raises questions about what it is we design in the interaction design process. What does it mean to shift focus from the things themselves to the acts that define them in use? [1].

Their question may be rephrased: What does it mean to shift focus from the scores themselves to the act of performing them? The products and things mentioned most likely imply technologies such as computers, cars, and domestic appliances. In this study, scores are regarded as products as well, with a focus on relations. That is, with reference to game design studies: “With interaction we mean both the interaction between players playing a game and the interaction between players and the game” [2].

Copyright: © 2024 First author et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ Interaction Design Foundation <https://www.interaction-design.org/literature/article/what-is-interaction-design>

This concerns the given and allowed interpretative freedom and the constraints afforded by a score, or simply how to interact with a given score. Therefore, interaction with a score happens *before* as well as *during* the performance. This can be described as two modes of music making, namely *design time* and *play time*; concepts that will be used in this text:

Design time, on the one hand, is outside time activity that deals with articulation and application of ideas and knowledge. On the other hand, play time is about real-time activity where interaction with the environment, embodied knowledge, and the present are at the forefront. [3].

This means, for a performer, that analysis and preparation intended for the performance of a score take place at design time, whereas the performance takes place at play time. Subsequently, the appearance of the music does not take place until it is performed, as Hallnäs and Redström assert:

In music the flow of tones in time is material that builds the actual appearance of a piece of music. When composing music, in contrast to ‘real time’ improvisation, we certainly work with this material but only in a sort of abstract manner – the actual compositional traces left are written instructions on how to perform the music, how to build its appearance. To set up these instructions we work with music technology helping us to work with a material that in a certain sense is absent most of the time. This makes it different from wood, paper etc. [4].

This view is shared by musicologist Christopher Small:

Properly understood, then, all art is performance act, which is to say that it is first and foremost activity. It is the act of art, the act of creating, of exhibiting, of performing, of viewing, of dancing, of wearing, of carrying in procession, of eating, of smelling, or of screaming that is important, not the created object. [5].

To conclude: a score contains unique information for how to perform a particular piece of music; musicians interact with the score, but the music only exists when it is played and/or heard in real time. In a later section, a short taxonomy regarding types of score–performer interaction will be proposed and defined.

Score vs Work

A musical work, in this context, is a collection of elements with a set of relations to each other that creates a unique sonic identity when performed. Elements can be many things, such as notes, rhythms, phrases, shapes, grooves, harmonic structures, sonic textures, timbres, as well as behaviors and interaction patterns that emerge. In other words, the given elements appear sequentially and/or simultaneously in an identifiable way in consecutive performances. A score, subsequently, is a set of instructions that are used to give the performers necessary information on how to perform a certain piece of music with preserved identity; it may range from a huge orchestra partiture to simple graphical shapes, as well as verbal or oral instructions. A score, however, is not the music, and according to Small:

A score, of course, is not a musical work. It is not even the representation of it. It is a set of coded instructions that, when properly carried out, will enable performers not only to make sounds in a specific combination, called a musical work, but also to repeat that combination as many times as they desire. Players and listeners learn to recognize that combination as a unity and to give it a name, which may be symphony no 5 in C minor [6].

Moreover, a written score is not even necessary to preserve the identity of a musical work in time and space. Here we are thinking of, e.g., the so-called “head-arrangement” used by American big bands in the 30s, notably Count Basie and Duke Ellington, i.e., pre-agreed and memorized arrangements, which were never written down.

One may say that the concept of a score started with Guido d’Arezzo, (ca 990–1050) when he invented the modern western musical notation system that eventually replaced the neumes that were in use at the time. One reason was to unify the performances of hymns in the Catholic Church, and to facilitate the learning of new material. However, one effect of Guido’s invention was the possibility to create and treat music at design time, e.g., to zoom in and out from details to overall form. Until the beginning of the 20th century, a score most likely primarily contained information of pitch and durations, combined with a set of established indicators of dynamics and articulation, and agreed terms for certain tempi and characters, like *andante* or *allegro*, together prescribing the way to perform a particular piece of music regarding phrasing, tempo, etc. Since this is well known, we will not pursue it further here.

Interaction

One aspect of a score is how much freedom performers are allowed. The short answer is that it depends. It depends, for example, on genre, style, the composer, tradition, and type of work. The range extends from music that demands the highest possible accuracy regarding, e.g., pitches, rhythm, and dynamics, to open-work scores that rely on and even force the player to make up her own decisions. One composer worth mentioning in this context is John Cage (1912–1992). In the chapter *Indeterminacy in Silence* [7], Cage discusses the terms determinate and indeterminate respectively, to distinguish between what is fixed and what is not. He exemplifies with two compositions:

The Klavierstück XI by Karlheinz Stockhausen is an example. *The Art of the Fugue* by Johann Sebastian Bach is an example. In *The Art of the Fugue*, structure, which is the division of the whole in parts; method, which is the note-to-note procedure; and form, which is the expressive content, the morphology of the continuity are all determined. Frequency and duration characteristic of the material are also determined. Timbre and amplitude characteristics of the material, by not being given, are indeterminate. This indeterminacy brings about the possibility of a unique overtone structure and decibel range for each performance of *The Art of the Fugue*. In the case of *Klavierstück XI*, all the characteristics of the material are determined, and so too is the note-to-note procedure, the method. The division of the whole into parts, the structure, is determinate. The sequence of these parts, however, is indeterminate, bringing about the

possibility of a unique form, which is to say a unique morphology of the continuity, a unique expressive content, for each performance. [8].

According to Cage, the Bach player is allowed to “color” the music, while in the Stockhausen example the performer rather has freedom, and are even obliged, to make choices with respect to the morphology of the continuity. More than 200 years separates Bach and Stockhausen, and in the rear-view mirror a trend appears towards more consciousness of the role of the score with respect to performance. Moreover, we cannot know exactly how Bach’s music was performed at his time. Rather, as artist Jeff Koons claims: “The past does not influence me; I influence the past” [9]. As the composer Alvin Lucier asserts about John Cage: “What have Cage’s pieces to do with self-expression? Nothing. They’ve got everything to do with discovery.” [10]. In *The Poetics of the Open Work* from 1959, Italian author Umberto Eco explains the open work by citing the composer Henri Pousseur: “[it] is not so much a musical composition as a field of possibilities, an explicit invitation to exercise choice” [11].

One consequence of the invention of the modern western musical notation was the possibility to make advanced treatments of musical material in the compositional process at design time. In turn, this development did increase the demands placed on the musicians, and the scores became less open for interpretative freedom at the performer’s own discretion. This development accelerated with the 1950’s serialism and its successors, where every musical element became constrained by compositional rules and, accordingly, the musicians became more of “sound-producing machines” whose role was to give sound to the composer’s visions and ideas, rather than self-expression. In serial music the scores contained very complex instructions, which made it difficult to judge whether a performance really did justice to the work. From when he was a music student, Nilsson remembers the Swedish modernist composer Jan W. Morthenson (1940–2019), who at a seminar in the 1980’s at the School of Music in Gothenburg voiced an extreme position: “we don’t need to listen to the music; it is enough to read the score.”

A conceptual idea that has been used in various ways in post-World War II avant-garde and experimental musical practices is to give performers unsolvable or contradictory tasks. John Cage employed the former strategy, inspired by Zen-Buddhist practices such as its short Koan forms, which most often are contradictory in content. Also worth mentioning is the “New Complexity” tradition, notably British composer Brian Ferneyhough who deliberately makes his scores partly unplayable, in order to consistently overstep the bounds of the human possible and consequently mistakes and erroneous playing become an intrinsic part of the performance. Another example is Miles Davis who routinely gave his musicians contradictory instructions. One reason for these strategies is to force performers to think and reflect to make up their own conclusions, and another is to change playing intention, from expressing something to solve a problem. All in all, the role and the function of a score have changed and developed since its inception by Guido d’Arezzo.

Interaction Types in Open Works

In an open work score, the task for the performer is not primarily to sonify the composer’s ideas, but rather to discover inherent possibilities in the score and to make choices. Moreover, an open work offers freedom to the performer but is still constrained to material and procedures provided in the score; the composer-designer purposely leaves certain decisions to the performer to solve. Out of Cage’s concepts of determinate and indeterminate in the examples above, two types of interaction emerge, which may be called *implied interaction*, and *commanded interaction* respectively. *Implied interaction* refers to a traditional interpretation process, for example in performing music of Bach. This means, when interpreting a score determinate in form and structure, a performer must make many decisions about parameters such as phrasing, tempi, dynamics, and timbre. The interpretation of the work takes place at design time; the player interacts with the score. *Commanded interaction* is when the score is indeterminate with respect to form and/or content, and explicitly commands a performer to interact, e.g., to choose the sequence of modules in the Stockhausen example. In addition, subgroups of these types may appear, such as *performers’ interaction*, which can be implied or commanded. *Emergent interaction* is a subgroup of implied interaction, which can be described as when characteristic interaction patterns emerge spontaneously, with no explicit instruction to do so.

APPLICATIONS AND ANALYSIS

In this section a selection of open works is presented and analyzed with respect to the types of interaction described above. Generally, the idea of the open work is attributed to the American composer Earle Brown (1926–2002). Compositions such as *Preliminary Notes* were inspired by the mobile sculptures by Alexander Calder (1898–1976). The shapes of the elements are fixed, but at the same time they are shifting. They show a stable identity, while in permanent variation. In *Preliminary Notes* the performer must choose the sequence of pre-composed parts. This is an example of commanded interaction regarding overall form, whereas the interpretation of individual parts is implied interaction.

One seminal work in this vein is Christian Wolff’s *For 1, 2 or 3 People* from 1963 (Wolff, 1963). Looking back, this composition is a significant milestone in the development of what a score may be. The score contains independent cells that Wolff called *ratio neumes*, which eventually became characteristic of Wolff’s score style, in which each action has a numeric ratio, according to Hicks and Asplund [12].

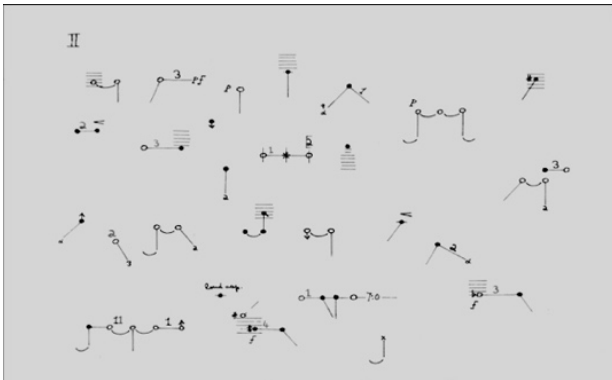


Figure 1: For 1, 2, or 3 People

For performance, the work prescribes a web of commanded interactions between the players, as well as between the players and the score. The first number in the ratio refers to a duration in seconds, whereas the second refers to how many of the given group of pitches (or pitches chosen ad lib by the performer). As Alvin Lucier asserts, one must first learn the vocabulary:

Black notes are variously short, up to about a second; white notes are any length. This may be confusing because in conventional notations black notes are usually short, white notes long. When you see a white note, you must remember that it's not necessarily long and may be determined by the requirements of coordination. Notes with stems as sixteenth notes are very short. A diagonal line toward a note indicates a sound played directly after a preceding sound. You coordinate with sound that came before. A diagonal line going away from a sound indicates that that note must be followed directly by another. [13].

Obviously, the commanded interaction is very complex, and in practice very difficult, or almost impossible, to pursue. If one fails, one must start over, which happens frequently, and therefore the identity of the piece, unintentionally, is rather a series of restarts.

Another example of a work in this spirit is *Treatise* by Cornelius Cardew (1965-67), a score containing 196 pages of abstract graphical elements, such as circles, squares, triangles and lines. At the outset Cardew did not provide any information of how to interpret the score, apart from two note stems at the bottom of each page, which implied traditional notation. Over the years, however, an interpretation tradition has evolved. Most often there are three performers; one plays the so-called lifeline in the middle, one the upper part and one the lower part. Eventually, in 1967 Cardew published *The Treatise Handbook* [14] that presents and discusses his view from experiences of performing the piece. One of the first performers of *Treatise*, together with Cardew, was British pianist John Tilbury, who in a personal conversation in 2013 explained and showed his interpretation, which gave us a clue of Cardew's ideas at the outset. Tilbury simply used the note stems at the bottom to notate what to play, and more than 50 years later he still uses it.

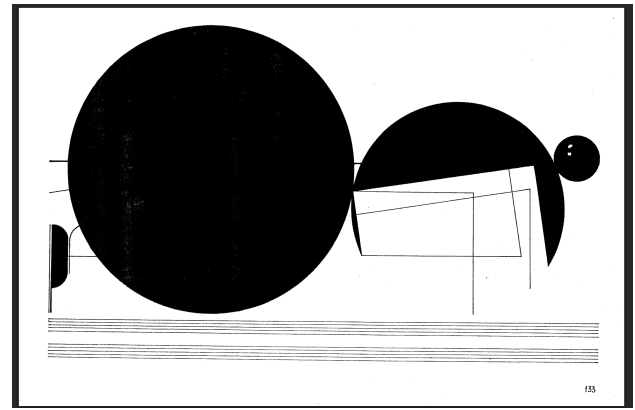


Figure 2: One page of Cornelius Cardew's *Treatise*

More recent performances do not have any use for conventional notation, particularly if performed with non-pitched electronic instruments. In our experience from performing *Treatise*, a great deal of the score-human interaction, however implicit, happens at design time. That is, a kind of negotiation between participating players is necessary, particularly regarding the mapping of graphic shapes to sound and playing behavior. However, what turns out to be essential, in our opinion, is to be consistent in one's interpretation of the different graphic elements. By doing so, the identity is preserved, and it also facilitates the interaction at play time to synchronize the ensemble.

Another piece by Cardew worth mentioning is *The Great Learning* from 1969. This suite of compositions contains seven parts and are based on Confucius book with the same name, which similarly has seven chapters. Each part is different from the others, but they all command interaction by the performers in different ways. Let us briefly look at *Chapter 6*. This part consists of 15 chunks of text, where each chunk starts with text fragments by Confucius in block letters, followed by a playing instruction. Performers should play the sequence of chunks in their own tempo, but be careful with perceived general pauses, and with synchronized events. The first chunks read as follows:

FROM Make or hear an isolated sound and hear out the general pause. Then a set of four sounds, the first one synchronized.

THE EMPEROR A pair of sounds, then a pair of optional sounds. [15].

This piece is first and foremost about commanded interaction, which leads to emergent interaction between the players. Written running instruction is straightforward, and additional performance notes explain sound production, such as optional, isolated and synchronized sounds. The difficulty is to synchronize sounds with another player; else you get stuck and must wait and search. To play this piece is to play a game; the score is played at each performer's individual speed, and therefore will finish at different times, and when finished, the player must leave the stage. At the end there is only one player left: the winner. Therefore, a possible tactic may be to hide the next sound, which makes it difficult for someone to synchronize, and they are forced to wait. Or sometimes two trapped players help each other to synchronize to be able to continue.

INTERACTIVE WORKS

In this section we present two oeuvres: *Chasing Chords* by Per Anders Nilsson and *The Bucket System*, a collaborative piece by Dahlstedt, Nilsson, Tim Perkis and Gino Robair, developed within an artistic research project. The common denominator is that they both command interaction and decision-making by the performers, else nothing will happen.

Chasing Chords

Chasing Chords [16] is a simple musical game that is inspired by first-person shooter video games, implemented in Cycling74 Max/MSP. In essence, it asks a drummer and a piano player to guess and to sync with sounds randomly generated by the software: a sampled bass drum and a piano chord respectively. The given task, however, to play in sync with the randomly generated events, is impossible to solve. The intention to do so, however, demands a certain attention from the participant players. In turn, this activity creates a unique quasi-periodic groove mediated by the software, and randomly generated events, in a mix of live playing and pre-recorded sampled sounds. The emergent groove is, in essence, the sonic identity of the piece, which is preserved between consecutive performances and with different players.

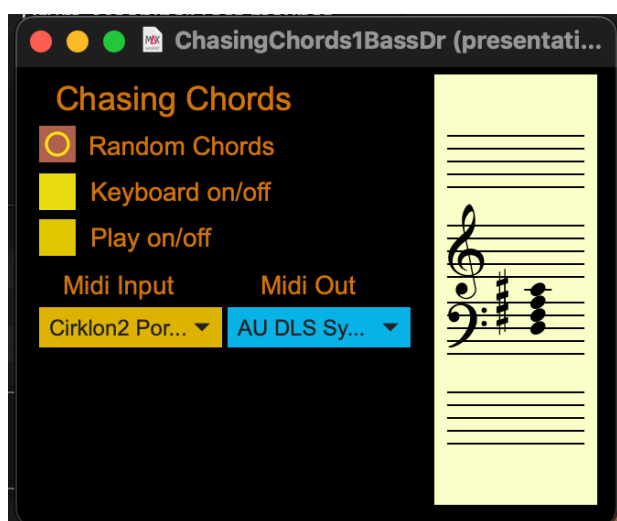


Figure 3: The Chasing Chord interface

This piece is, according to Cage's nomenclature, determinate in structure, and indeterminate in performance. Moreover, there is no need for interpretative work before performance, at design time. Instead, the players just need to understand the task.

The Bucket System family

The Bucket System (Dahlstedt, Nilsson, Perkis, Robair 2015-2025) allows for a new kind of on-stage compositional/improvisation interaction. In essence, it is a signaling system that mediates players' instructions for

playing behavior and/or relations to fellow players. At the outset it is a system for computer-mediated ensemble improvisation, and its credo states that players are composers and performers at the same time. It emerges from a tradition of structured free ensemble improvisation, such as *conduction* and *comprovisation*, and is also influenced by post-World War II experimental, free improvisation music practices, as well as by theatre improvisation techniques. Before playing, participants need to negotiate and define what the four possible signals mean. In this sense the Bucket System is, in line with Cardew's Treatise, an open structure whose content must be defined as part of the interaction with the score at design time. During play, at any moment, any player can submit configurations of signals to a virtual bucket, a storage of potential future configurations of roles for the players. At random times (regulated by some constraints, e.g., minimum and maximum waiting times), a random configuration is "picked up", i.e., selected, so that each player suddenly receives a new signal that (s)he is forced to follow. At its first inception in 2015 [17], implemented using a McMillen *QuNeo* interface, the signals consisted of four led light states: slow and fast blinking light, steady light, and no light. The signals were given on a set of LED buttons that represent a geographical map of the four players, so each player is assigned one state, and each musician can see all four signals. A later version, *Bucket 2*, uses touch screens (as an app on tablets, phones or touchscreen computers) and allows for up to 12 players, using four colors as signals: red, yellow, blue and dark grey.



Figure 4: The Bucket 2 interface

In the Bucket System, an interpretation denotes a set of assigned meanings to the four signals. Here are two examples of interpretations out of the many that have been used with Bucket:

Metaphorical: Red = Busy, Yellow = Simple, Blue = Support².

This interpretation asks the players to focus on their own playing, except for blue, which demands active listening.

² Metaphorical comes from the original Bucket, where the signals consisted of blinking LED lights, where faster blinking it itself might induce more energetic playing.

No other information is given. The musical outcome is a result from the interaction that emerges when participating players fulfill their sequence of given interpretations.

Behavioral: Red = Solo, Yellow = Interact, Blue = Vacillate.

Here the players are forced to listen to the other participants and adjust their playing in accordance with the present task. It is about distribution of roles, and “vacillate” affords a certain degree of freedom, but also introduces ambiguity.

Metaphor and Behavioral are arbitrary names, derived from its basic type of interpretation, to easily discern different signal assignments, or interpretations as we call them. In practice, during a performance, the players may change interpretation, either in a break, or during performance.

Yet another implementation of the Bucket System was developed using color-coded dice. Basically, it is the same concept regarding content, using similar interpretations as presented above. The difference is in the procedures of changing tasks for individual players. In the dice interpretations, any player can at any moment command any other player (or all players) to throw their dice, signaled through hand gestures.

Let us conclude with a recent implementation of ours within the Bucket System family. It is a game board of 55 x 55 cm, made for a performance with four musicians and two dancers. The game board is placed on the stage floor, visible to both the players and the audience. This version contains four basic colors: yellow, red, blue and green, to signal performers' roles, with the addition of black and white as special modifiers representing the dancers. For example, the metaphor interpretation here is given as: Red = Busy, Yellow = Simple, Blue = Extended, Green = Intermittent. Its procedure assigning signals involves physical action by the performers: players are not allowed to change their own interpretation, but are free, and even obliged, to add, remove, or change color for the other players squares and place the dancer's markers. For the two dancers the black and the white cards (symbolizing each dancer) were assigned in such a way that their placement (on top of a musician's field, or across two or more musicians' fields), shows which musician(s) they should interact with, applying interaction patterns from the underlying musicians' color signals. This is a typical example of a partly insolvable task.



Figure 4: Bucket as Floor Game

Comments

The following features are common for all members of the Bucket System family:

- 1) There is no assigned leader or conductor, all participants have a similar role;
- 2) participants must negotiate and agree upon interpretation of the different signals/colors before play;
- 3) participants must follow the current individual signal during play;
- 4) participants must “fill” the bucket/gameboard with new constellations of interpretations, while playing.

In the different versions, procedures of selection of next constellations, as well as how new constellations are introduced differs. This means that the performers, during play, must fulfill two tasks: firstly, to play and interact according to actual color, and secondly to figure out the continuation of the music, with new constellations, else the actual constellation will get stuck, however within the limits of the agreed interpretation.

CONCLUSION

Instead of thinking that a composition, and its score, is a product, we instead envision the score or instruction set as follows: “it is not so much a musical composition as a field of possibilities, an explicit invitation to exercise choice” [18]. As composers, or musical interaction designers, we invite participant players to play with our pieces (or systems), to investigate inherent possibilities in the given work. However, in our opinion, it is naïve to claim that interactive pieces are not shaped by the biases of their designer, about what to expect regarding the interaction at play time and regarding musical outcome. From time to time, participating players will do something unexpected that challenges the intention and expected outcome. As the designer, you may accept and embrace this and think it is interesting and realize that there are hidden possibilities that you were not aware of. But in certain cases, playing and subsequently the outcome becomes boring, when players exhaust intrinsic possibilities too quickly. Or, that the playing takes a direction that goes too far from the original idea, e.g., because players fall back into their own habitual ways of playing. In such cases, one must go back, and at design time re-design the set of rules.

To conclude: in an interaction-based work you invite players to be co-creators within the provided game plan. One must be prepared, and even encourage the players, to try to challenge the work, but still obey the overall constraints, and preserve the intended identity. Most likely, before play-time activity, discussions and negotiations at design time will occur. In our experience, it is difficult to create a score, or game-like instructions, which covers all aspects. The driving force to make interaction work is not to create perfect rule systems, but rather to involve participating players in a creative act by providing creative constraints. It should be fun to play!

REFERENCES

1. Hallnäs, Lars, Redström, Johan 2006, *Interaction Design: Foundations*, p. 22
2. Björk, Lundgren and Holopainen, *Game Design Patterns*. 2003, p.4
3. Nilsson, Per Anders, 2011, *A Field of Possibilities – Designing and Playing with Digital Musical Instruments*, Art Monitor, Univ. of Gothenburg, p. 27
4. Hallnäs, Redström, 2006, *Interaction Design: Foundations*, p. 108-109
5. Small, Christopher, 1998, *Musicking: The meanings of performing and listening*, Univ. Press of New England. p.108
6. Small, 1998, p. 112
7. Cage, John, 1961, *Silence: Lectures and Writings*, Wesleyan University, p.35-40
8. Cage, 1961, *Silence*, p.35
9. Cage, 1961, *Silence*, p.67
10. Lucier, 2012, *Music 109*, Wesleyan University.
11. Henri Pousseur as cited in Eco, Umberto 1959, 'The Poetics of the Open Work', in *Audio Culture: Readings in modern music*, 2004, ed. Cox, C., and Warner, D. Continuum, New York, pp. 167–175. p. 168
12. Hicks, M, Asplund, Christian, 2012, *Christian Wolff*, University of Illinois, p.24
13. Lucier, 2012, p.46, 57
14. Cardew, Cornelius, 1967, *Treatise Handbook*, The Gallery Upstairs Press
15. Cardew, Corenlius, 1969, *The Great Learning*, score. Edition Peters
16. Nilsson, Per Anders, Natural Artefacts, 2014, *California Connection*, CD. LJ-Records LJCD-5258
17. Dahlstedt, Palle, Nilsson, Per Anders, Robair, Gino 2015, *The Bucket System*, Proceedings of NIME 2015, Baton Rouge, LA, USA.
18. Pousseur in Eco, 1959, p. 168.

VISUAL SCORES AS LISTENING-MEDIATED NOTATION: A PRACTICE-BASED CASE STUDY

Grgur Savić

Independent Researcher
grgasavic@gmail.com

ABSTRACT

This paper presents a practice-based analysis of a time-based visual score, accompanied by extensive performance notes specifying symbol execution. The digitally animated notation employs relational, temporal, and gestural cues organized along a moving horizontal timeline, regulating entrances, transitions, and pauses without prescribing fixed pitches or durations. Performers translate visual information into sound through attentive listening and ensemble negotiation. Drawing on rehearsal observations and performance analysis, the study examines how this approach fosters interpretive flexibility, coordination, and collective awareness. The findings highlight complementary perspectives to audio-centric notation, while also outlining pedagogical implications for rehearsal processes, ensemble training, and the development of listening-based performance practices.

1. INTRODUCTION

The present work consists of a time-based visual system in which symbols and graphical elements are accompanied by an extensive set of performance notes specifying their execution [7]. Rather than encoding fixed pitches or conventional rhythmic notation, the score employs relational, temporal, and gestural cues that unfold along a moving horizontal timeline, indicated by a continuous marker that regulates entrances, transitions, and moments of silence. Performers interpret these cues in real time, translating visual information into sound while continuously negotiating their actions through attentive listening and embodied interaction [9]. In this sense, the work does not function as a prescriptive representation of sonic results but as a mediating frame-

work that coordinates temporal awareness, collective listening, and interpretive responsibility within the ensemble.

This approach relates to existing practices in graphic and animated notation, particularly scrolling and time-based visual scores that structure performance through moving visual cues rather than fixed symbolic encoding. Previous work has explored similar relationships between visual representation, temporal coordination, and performer interpretation in experimental and digital notation systems [10]–[15]. Within the context of TENOR research, particular attention has been given to animated notation, performer interaction, and embodied interpretation, emphasizing the role of perception and real-time negotiation in shaping musical outcomes. These perspectives situate the present work within a broader discourse on non-traditional and technology-mediated notation practices.

This paper examines how a time-based visual score mediates collective listening and interpretation in performance, and how this perspective contributes to ongoing discussions on visual and audio-centric approaches to contemporary notation.

2. PERFORMANCE NOTES

The notation examined here is accompanied by twenty-seven pages of performance notes that define the execution of each visual symbol, the relationship between graphical elements and sound production, and the temporal logic of the piece [7]. Rather than prescribing fixed sonic outcomes, these notes establish a shared interpretive framework that ensures consistency of execution while allowing performers to respond to situational listening and ensemble interaction. The moving horizontal timeline provides a shared temporal reference, while the visual symbols and written instructions frame interpretive decision-making without prescribing fixed sonic results; in this sense, the notation

Copyright: © 2026 Grgur Savić. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

functions less as a representational system and more as a protocol for real-time musical action [7], [9].

As shown in Figures 1–3, the notation employs distinct visual elements to differentiate modes of sound production and performer action.

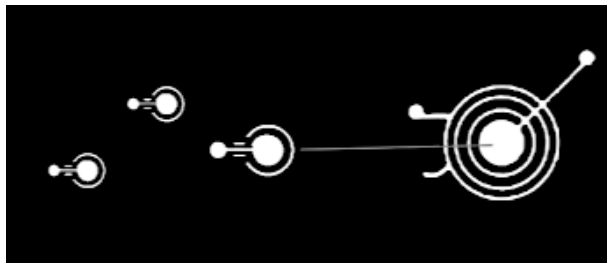


Figure 1. Symbol indicating high-frequency electronic textures (e.g., bleeps, sonar-like signals, and metallic timbres) with sustained background tones. The gesture is performed with a gradual crescendo, guiding a soft entry and increasing intensity over time.

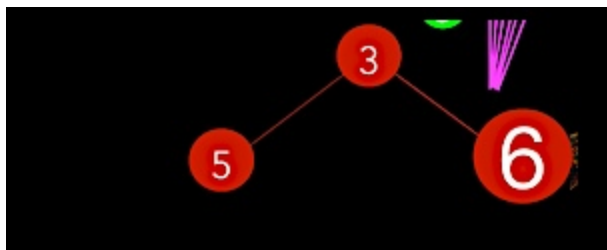


Figure 2. Symbol representing flexible rhythmic groupings (e.g., quintuplets, triplets, sextuplets) coordinated with the moving timeline. Connecting lines enable extension and repetition, allowing performers to generate short melodic patterns within chosen subdivisions.

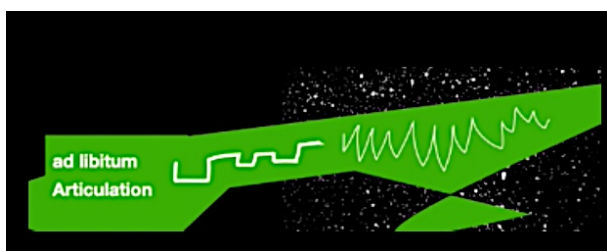


Figure 3. Symbol for ad libitum articulation within the moving timeline, guided by attentive listening and ensemble blending. The inner line indicates legato continuity with flexible pitch selection, while increased visual density reflects greater intensity and expressive character.

For example, certain symbols indicate modes of sound production rather than fixed pitches: a curved line may signal gradual timbral transformation, while dense clustered shapes suggest textural activity or increased sonic intensity. Color coding differentiates functional roles, such as sustained tones, percussive gestures, or

transitional events, allowing performers to negotiate their actions in real time [11], [12].

These elements do not prescribe exact sonic results but establish a shared interpretive vocabulary for coordinated interpretation.

Such a structure encourages a creative approach to performance grounded in musical decision-making that connects auditory experience with visual perception. It establishes relationships between pitch, articulation, dynamics, and duration while fostering both coherence and perceptual tension [1], [2]. This framework combines precisely defined visual elements with open-ended interpretive zones, forming what can be described as a framed notational system: a structure that provides direction while inviting exploration.

3. SCORE EXAMPLES

Excerpt from the digital-animated visual score showing the time-based structure of the piece. A continuous horizontal marker progresses from left to right, regulating entrances, transitions, and pauses. Graphical symbols are interpreted according to the accompanying performance notes and do not encode fixed pitches or durations. The excerpt illustrates how temporal coordination and performer interaction are guided visually while sonic outcomes remain listening-dependent [7].

4. CASE STUDY

4.1 Performance Context

The system was realized in a professional chamber music performance context involving wind players and live electronics. The program combined twentieth-century classical works [3]–[6] with contemporary graphic scores [1], [2], situating the present work [7] alongside both traditional and experimental notation practices. The ensemble consisted of five musicians: flute, oboe, clarinet, bassoon, and French horn, supplemented by live electronics using a modular synthesizer. This configuration enabled a direct comparison between conventional and experimental approaches within a professional ensemble setting.

4.2 Preparation Process

The ensemble met for three preparatory sessions of approximately four hours each prior to the concert. These sessions were shared between classical repertoire and graphic notation, allowing one to two rehearsals of each graphic score.

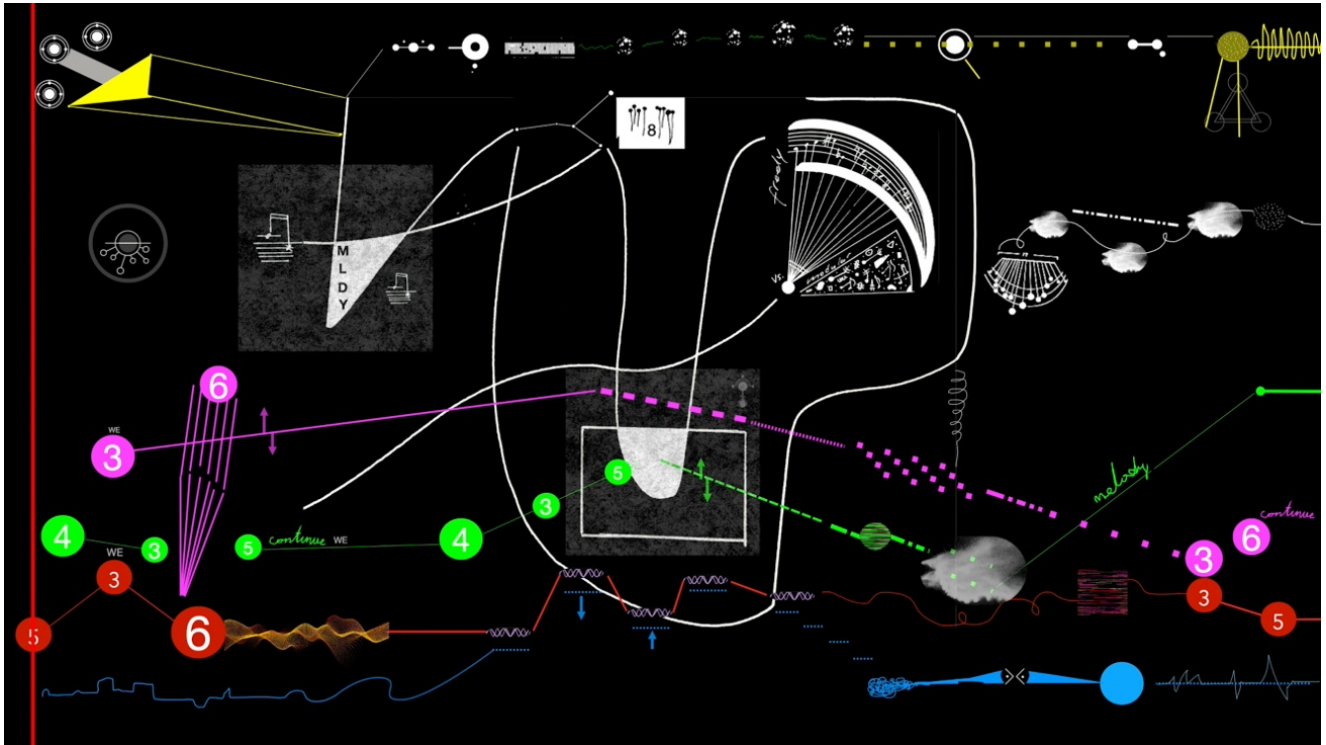


Figure 4. Excerpt from the digital animated visual score *Eternal Forces of Savigny's Prime*, showing the moving timeline, symbolic layers, and visual density used for temporal coordination.

The present score was accompanied by twenty-seven pages of detailed performance notes, covering 105 abstract symbols, temporal logic, and instrument-specific instructions [7]. The composer conducted one rehearsal to provide temporal guidance and dynamic emphasis, supporting ensemble cohesion without imposing fixed sonic outcomes.

4.3 Listening and Interaction

During rehearsals, performers displayed varied strategies in engaging with the score. Some relied heavily on the performance notes, carefully interpreting symbols and timings; others interacted more intuitively with the visual elements, using them as cues for improvisational decisions. As shown in Figure 5, performers followed the moving timeline as a shared temporal reference, illustrating the balance between structural guidance and interpretive freedom [7], [9]. This highlighted the role of performer experience, personality, and familiarity with graphic notation in shaping interpretive outcomes.

For instance, performers frequently relied on auditory cues from one another to adjust timing and dynamics when visual information alone was insufficient. In several passages, entrances were coordinated not through strict adherence to the timeline but by responding to emerging textures within the ensemble. This contrasts with conventional notation, where synchronization is typically derived from fixed rhythmic structures, as here

coordination emerges through continuous listening and real-time adjustment between performers [11].

As shown in Figures 6–8, the visual score functions not only as a temporal guide, but also as a spatial and perceptual framework, shaping performer orientation, ensemble interaction, and collective listening during performance.

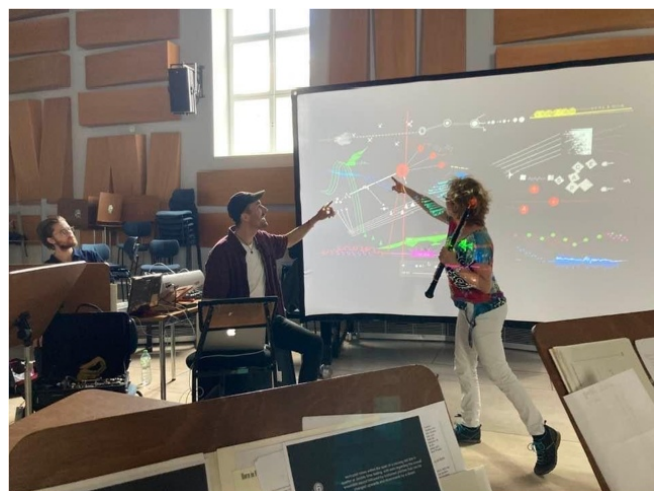


Figure 5. First rehearsal at Haus des Rundfunks, with the composer providing temporal and dynamic guidance. Performers reference both the visual score and accompanying performance notes.



Figure 6. Performance configuration at Kühlhaus Berlin, showing the ensemble, projected score, and audience arrangement. The visual score functions as a shared temporal and spatial reference for the performers.



Figure 7. Electronics setup and projected score in a circular viewing configuration. This spatial arrangement introduced an additional layer of visual orientation, affecting performer interaction, perception, and ensemble listening.

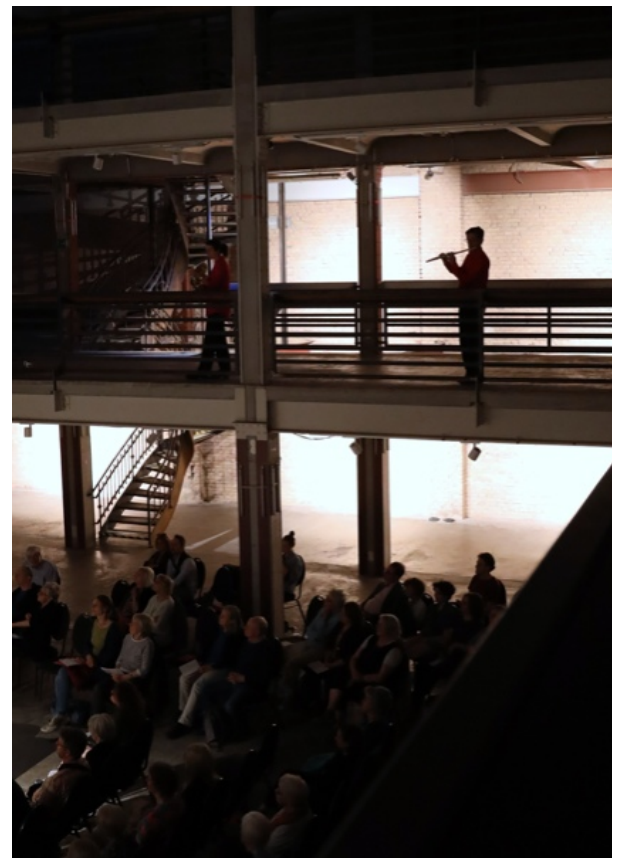


Figure 8. Spatial performance configuration at Kühlhaus Berlin. Ensemble members perform while viewing the projected visual score from the second floor in a 360-degree formation, integrating performer positioning, visual orientation, and collective listening.

Despite limited rehearsal time, the ensemble maintained a high degree of temporal precision in relation to the moving timeline. Sections conducted by the composer demonstrated increased dynamic contrast and clarity, whereas unassisted run-throughs required more negotiation among performers. These observations indicate that the score operates as a listening-mediated system, fostering collective attention and coordination even in constrained preparation contexts [7], [9].

In this sense, the score functions less as a directive system and more as a framework that distributes responsibility for temporal and sonic coordination across the ensemble.

4.4 Reflections

The case study prompted a reassessment of how visual scores mediate between guidance, interpretive freedom, and performer engagement. One key insight was that the effectiveness of such notation systems depends not only on their internal logic but also on performers' familiarity with contemporary practices and their willingness to engage with detailed symbolic frameworks. While some musicians responded intuitively to the visual material, others required more explicit clarification of symbol execution, reflecting differing interpretive approaches shaped by background and experience.

Extensive performance notes proved both beneficial and challenging. On one hand, detailed documentation supported consistency, temporal alignment, and ensemble cohesion. On the other, the volume of information required significant preparation time, occasionally leading to uncertainty or uneven engagement among performers. This highlighted a central tension in the practice: balancing sufficient guidance to ensure coordinated realization while preserving space for creativity, improvisation, and listening-based decision-making.

Adaptive strategies emerged in response to these challenges. In situations of interpretive uncertainty or temporal instability, conducting or gestural cueing reinforced shared tempo, dynamic contrast, and ensemble focus without imposing fixed sonic outcomes. These interventions functioned as temporary supports, maintaining the score's listening-mediated character while addressing practical performance needs.

Comparisons between ensembles further clarified the role of professional training and rehearsal culture. Musicians accustomed to regular rehearsal schedules and diverse tempo demands demonstrated greater temporal precision and confidence in navigating the moving timeline, while

others required more negotiation and adjustment. These observations suggest that the same visual system can yield different outcomes depending on context, emphasizing the importance of adaptability within the compositional framework.

Ultimately, the practice underscores the value of designing visual scores as flexible yet coherent systems that articulate a clear symbolic language while remaining responsive to performer interpretation and situational listening. By integrating handwritten and digital elements, the notation establishes a structured but open-ended environment in which performers actively shape musical outcomes in real time [7], [9], [10].

5. CONCLUSION

This study demonstrates, through rehearsal observations and performance analysis, that visual scores accompanied by performance notes can function as effective listening-mediated notation systems [7], [9]. In relation to contemporary audio-score practices, this work emphasizes the continued relevance of visual mediation and performer agency within experimental notation [9], [10]. Rather than opposing audio-centric approaches, the score offers a complementary perspective in which visual structures guide listening, interaction, and temporal awareness.

Moreover, the score has pedagogical implications for contemporary music practice by foregrounding listening as an active, collective, and interpretive process [7], [9]. Performers negotiate visual cues through auditory attention, situational awareness, and ensemble interaction, cultivating interpretive responsibility and collaborative sensitivity. In this way, the notation functions not only as a performative framework but also as a reflective tool for examining relationships between notation, perception, and collective musical practice.

Visual scores, understood in this manner, contribute both to artistic production and to broader discussions on the interplay between visual and audio-centric approaches to contemporary music notation [9], [10].

6. REFERENCES

- [1] B. Bacon, *Gesture Contours I–III*, score, 2024.
- [2] J. Kern, *Linear Cycles 1*, score, 2024.
- [3] A. von Zemlinsky, *Humoreske*, score, 1939.
- [4] D. Milhaud, *La Cheminée du roi René*, score, 1939.

- [5] K. Weill, *Lieder aus Die Dreigroschenoper*, score, 1928.
- [6] P. Hindemith, *Kleine Kammermusik für fünf Bläser "Des kleinen Elektromusikers Lieblinge"*, score, 1922.
- [7] G. Savić, "Eternal Forces of Savigny's Prime," visual score and performance notes, 2024.
- [8] G. Savić, *The Art of Visual Music: Exploration of Graphic Notations in Compositional and Improvisational Contexts in Contemporary Music*, MA thesis, CATALYST – Institute for Creative Arts and Technology, 2024.
- [9] Personal communication, 2024.
- [10] C. Cardew, "Notation—Interpretation, etc.," *Tempo*, no. 58, pp. 21–33, 1961.
- [11] J. Hope, "From score to screen: Animated notation in contemporary music," *Contemporary Music Review*, vol. 36, no. 1–2, pp. 123–145, 2017.
- [12] T. Winkler, "The role of gesture in interactive music," in *Proceedings of the International Computer Music Conference (ICMC)*, 1995.
- [13] C. Roads, *The Computer Music Tutorial*. Cambridge, MA: MIT Press, 1996.
- [14] T. Magnusson, "Scoring with code: Composing with algorithmic notation," in *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR)*, 2019.
- [15] A. Nash, "Animated notation and the performer: Interaction, interpretation and embodiment," in *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR)*, 2018.

PREScriptive NOTATION AS PERFORMATIVE INTERFACE

Djordje Marković

PhD student, Academy of
Arts, University of Novi Sad,
Serbia
getmoll@gmail.com

ABSTRACT

This paper approaches musical notation not as a representational system for sound, but as a prescriptive interface that organizes instrumental action and conditions the emergence of musical events. Drawing on the performative turn in music, improvisation studies, and Gilles Deleuze's concept of virtuality, it argues that increased prescriptivity often intensifies rather than reduces interpretative openness. By examining prescriptive layers embedded within classical notation – such as articulation, ornamentation, and instrumental instructions – the paper challenges the opposition between descriptive and prescriptive notation. Prescriptivity is understood as operating below the level of sound, at the level of embodied action, allowing sonic outcomes to remain contingent and immanent to performance. The argument is grounded in the author's compositional practice, with examples from trumpet notation that employ minimally extended classical notation to prescribe complex actions while resisting stable sonic representation. The paper concludes by proposing prescriptive notation as a catalyst for musical *becoming* rather than a blueprint for musical objects.

1. INTRODUCTION: NOTATION BEYOND REPRESENTATION

Musical notation has traditionally been understood as a representational system whose primary function is to describe sound. In this view, the score operates as a visual encoding of musical parameters – most notably pitch, duration, and dynamics – serving as a stable intermediary between compositional intention and sonic realization. Such an understanding presupposes that the musical work exists, at least in principle, as an identifiable sound structure that can be adequately captured, transmitted, and reproduced through notation. This characterization should not be taken to imply that traditional notation lacks prescriptive or action-oriented dimensions. On the contrary, articulation, phrasing, and technical indications already encode forms of instrumental action. However,

within a representational paradigm, these elements typically remain subordinated to the specification of sonic outcomes, rather than constituting the primary object of notation. Although this model has proven historically resilient and remains foundational for large parts of the Western art music tradition, where concepts such as the “musical idea” or structural coherence continue to play a central role. The present argument does not seek to dismiss this framework, but to identify the conditions under which its underlying assumptions become insufficient.

However, this representational paradigm becomes increasingly problematic when confronted with musical practices that foreground performance, embodiment, and sonic instability. Improvisation, extended instrumental techniques, and process-oriented compositional approaches all expose a fundamental tension between notation understood as a description of sonic results and music understood as an event that unfolds through action. In these contexts, sound cannot be treated as a fixed object to be represented; rather, it emerges as a contingent outcome of situated interactions between performer, instrument, and environment.

This paper argues that musical notation should therefore be reconsidered not primarily as a representational medium, but as a performative interface: a system that prescribes actions, constrains possibilities, and organizes the conditions under which sound can occur. In this context, the notion of interface does not refer to a technological device, but to a relational structure that mediates between the performer's body, the instrument, and the unfolding sonic process. From this perspective, notation does not merely encode (pre-existing) musical objects; it actively participates in their production by shaping modes of performance, bodily engagement, and instrumental behavior.

Central to this argument is the distinction between descriptive and prescriptive aspects of notation. Descriptive notation aims to specify sonic outcomes – what should be heard – while prescriptive notation focuses on actions and procedures – what should be done. While this distinction is often presented as a modern opposition between traditional notation and experimental or graphic scores, this paper contends that prescriptivity has always been an integral, if frequently naturalized, component of Western art music notation. Articulation marks, technical instructions, ornaments, and performance conventions already operate prescriptively by directing bodily and

Copyright: © 2026 First author et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

instrumental actions without fully determining their sonic results.

Rather than treating prescriptive notation as a deviation from or supplement to an otherwise descriptive system, this study proposes to understand prescriptivity as a latent structural dimension of notation that becomes explicit under certain historical and aesthetic conditions. In contemporary music, particularly in practices that engage with improvisation and unstable sound production, this prescriptive dimension becomes central. Here, notation no longer aims to fix sound identities, but to articulate regimes of action whose sonic outcomes remain open and immanent to the act of performance.

This shift resonates with broader theoretical developments often described as the performative turn in music and the arts. Influenced by philosophical accounts that emphasize process, becoming, and event over representation and objecthood, these approaches reconceptualize music as something that happens rather than something that is. Within such a framework, the musical work is not a stable entity preceding its realization, but an emergent configuration that takes shape through performance.

The present paper situates prescriptive notation within this performative and process-oriented ontology of music. Drawing on concepts from Deleuze's philosophy of immanence and transcendental empiricism, it argues that prescriptive notation operates at the level of the virtual: it does not define actual sounds, but configures fields of potential actions and tendencies that are actualized differently in each performance. Precision at the level of action thus coexists with openness at the level of sonic result, challenging the assumption that notational determinacy necessarily entails sonic determinacy.

Methodologically, the paper combines theoretical reflection with analysis of compositional practice. Alongside a conceptual discussion of notation, improvisation, and musical ontology, it examines examples from the author's own work, focusing particularly on the use of minimally extended classical notation to prescribe instrumental actions – especially on the trumpet – while leaving the resulting sound partially indeterminate. These case studies serve not as illustrations of pre-existing theory, but as sites where theoretical and practical concerns intersect and inform one another.

By reframing musical notation as a performative interface rather than a representational code, this paper contributes to ongoing discussions in music theory, aesthetics, and notation studies concerning the ontology of musical works, the role of performance, and the limits of representation. It aims to show that prescriptive notation is not a retreat from precision, nor a mere strategy of indeterminacy, but a means of articulating music as an event whose identity resides in processes of action and becoming rather than in fixed sonic forms. While the Deleuzian framework provides a productive vocabulary for articulating these processes, the argument developed here does not depend exclusively on this philosophical model, and may also be understood in more practice-oriented or phenomenological terms.

2. THE PERFORMATIVE TURN IN MUSIC

The notion of a performative turn in music designates a broad shift in theoretical and artistic thinking away from the conception of music as a fixed object and toward an understanding of music as an event, a process, or an activity that unfolds in time. Rather than treating the musical work as an autonomous structure that precedes its realization, performative approaches emphasize the constitutive role of performance, embodiment, and situational context in the emergence of musical meaning.

Historically, Western art music has largely been grounded in an object-centered ontology, in which the musical work is identified with a notated structure whose identity remains stable across performances [1]. Performance, within this paradigm, is understood as the execution or interpretation of a pre-existing work, and notation functions as the primary guarantor of that work's identity [1, 6]. The performative turn challenges this hierarchy by questioning the assumption that the work exists independently of its enactment.

In this alternative perspective, performance is no longer a secondary act that realizes an already defined musical object, but a primary site where music comes into being [2]. Sound is not the reproduction of a pre-given form; it is the outcome of situated actions involving bodies, instruments, gestures, and acoustic conditions. Music thus appears less as an entity to be represented and more as a dynamic process of emergence.

Improvisation occupies a particularly important position within this shift. Beyond its status as a stylistic or genre-based practice, improvisation foregrounds the temporal contingency and irreversibility of musical action [3, 4]. In improvisatory contexts, musical decisions are made in real time, without the mediation of a fixed score, and the identity of the music is inseparable from the concrete circumstances of its production. What is crucial in this context is not simply spontaneity, but the irreversibility of the temporal process: once an action is performed, it cannot be retracted or fully pre-specified. This irreversibility challenges the assumption that musical form can be entirely determined in advance through representation. Improvisation therefore exemplifies a mode of musical thinking in which action precedes representation and sound emerges directly from embodied interaction.

This does not imply that improvisation is opposed to structure or constraint. On the contrary, improvisatory practices often rely on highly developed technical vocabularies and implicit rules. What distinguishes them is not the absence of organization, but the fact that organization is enacted rather than pre-inscribed. From the perspective of notation, improvisation exposes the limits of descriptive systems that presuppose stable sonic targets, highlighting instead the primacy of procedural knowledge and performative competence.

Philosophical accounts associated with process-oriented ontologies provide a conceptual framework for understanding this shift. According to the work of Gilles Deleuze, music can be approached not as a collection of fixed forms but as a field of becoming, where events

actualize virtual tendencies rather than realize predetermined identities [5]. The distinction between the virtual and the actual is particularly relevant here: the virtual is not unreal or abstract, but a structured field of potentialities that are actualized differently in each concrete situation.

Within such a framework, a musical work cannot be reduced to a set of sonic properties. Its identity resides instead in a configuration of forces, tendencies, and constraints that give rise to multiple possible realizations. Performance is not the execution of an ideal model, but the site where these virtual structures are actualized through material action. Music, in this sense, is fundamentally performative.

The performative turn thus destabilizes the traditional role of notation as a repository of musical identity. If music is understood as an event rather than an object, then notation can no longer be conceived solely as a means of representing sound. Instead, it must be understood in relation to the actions and processes through which sound emerges. This reconceptualization does not render notation obsolete; rather, it invites a rethinking of its function and scope [6].

From this perspective, the question is no longer how accurately notation can describe sound, but how it participates in shaping musical action. Notation becomes a tool for organizing performative conditions rather than a blueprint for sonic outcomes. This shift lays the conceptual groundwork for a prescriptive understanding of notation, in which the score specifies procedures, gestures, and constraints while leaving the sonic result open to the contingencies of performance.

By situating notation within the performative turn, this paper prepares the ground for a re-evaluation of the relationship between composition, improvisation, and musical ontology. The distinction proposed here should therefore not be understood as a strict opposition between notation and improvisation, but as a shift in emphasis from reversible representational structures toward irreversible processes of action unfolding in time. The following section builds on this framework by examining improvisation more closely as a practice that exposes the limits of descriptive notation and points toward alternative, prescriptive modes of musical inscription.

3. IMPROVISATION AND THE LIMITS OF DESCRIPTIVE NOTATION

Improvisation poses a fundamental challenge to musical notation understood primarily as a descriptive system. If notation aims to represent sound by specifying pitch, duration, and dynamic relationships in advance, improvisation exposes the limits of such representation by foregrounding musical action as a real-time, irreversible process. In improvisatory contexts, sound is not the realization of a pre-existing structure but the immediate outcome of decisions made in performance. The question that emerges is not how to notate an improvised result, but how notation can meaningfully relate to a practice in which musical identity is inseparable from action.

From the perspective of descriptive notation, improvisation appears as a problem of

underdetermination. The more music depends on contingent interactions between performer, instrument, and moment, the less adequately it can be captured by a system designed to encode stable sonic outcomes. Attempts to fully describe improvisation through traditional notation often result either in excessive abstraction – reducing music to schematic outlines – or in an illusion of control that obscures the performative reality of the practice [3].

Historically, this tension is already visible within the canon of Western art music itself. One of the most prominent examples is the solo cadenza in the classical concerto. Traditionally positioned near the end of the first movement, the cadenza marks a suspension of orchestral time and a transfer of compositional agency to the soloist. While formally integrated into the work, the cadenza was for a long period not fully notated, but rather indicated by a fermata or a conventional cue, leaving its realization to the performer's improvisatory skill.

The cadenza thus functions as a liminal space within an otherwise highly regulated notational system. It reveals that even within a predominantly descriptive framework, certain musical situations resist full inscription. What is left unnotated in the cadenza is not the sound itself, but the sequence of actions, gestures, and decisions through which sound comes into being. The eventual practice of composers writing out cadenzas in the nineteenth century can be understood as an attempt to reassert control over this indeterminate zone, effectively transforming an action-based practice into a descriptive object.

A similar tension characterizes the relationship between improvisation and extended instrumental techniques in contemporary music. Many such techniques – multiphonics, unstable timbres, noise-based sounds – cannot be reliably specified in terms of fixed sonic parameters. Their realization depends on fine-grained bodily actions, micro-adjustments, and instrument-specific behaviors that exceed the descriptive capacity of traditional notation. As a result, composers often resort to verbal instructions, diagrams, or hybrid notational strategies that shift the focus from sound description to action prescription.

This shift highlights a crucial point: improvisation does not simply fall outside notation; it reveals a different ontological layer of musical practice. Improvisation operates at the level of how sound is produced rather than what sound should result. From this perspective, the inadequacy of descriptive notation is not a technical limitation, but a conceptual one. Descriptive notation presupposes that sound can be abstracted from the actions that produce it, whereas improvisation insists on their inseparability.

The limits of descriptive notation thus become visible precisely where musical meaning is grounded in performative causality rather than in sonic identity. Improvisation foregrounds the fact that many musical parameters – timbre, instability, intensity, resistance – are not properties of sound objects but effects of action. These parameters cannot be fully captured by symbolic representation; they must be enacted.

This realization does not imply that improvisation necessitates the abandonment of notation altogether.

Rather, it invites a rethinking of notation's function. If notation is understood not as a mirror of sound but as a mediator of musical action, then its role in improvisatory contexts can be reconfigured. Instead of prescribing sonic results, notation can specify constraints, tendencies, or action-spaces within which improvisation unfolds [7, 8].

In this sense, improvisation and notation are not opposites but exist in a productive tension. Improvisation exposes the limits of descriptive notation, while notation, reconceived prescriptively, can articulate conditions that shape improvisatory behavior without eliminating its openness. This perspective paves the way for understanding prescriptive notation as a means of engaging with improvisation without reducing it to a fixed object.

At the same time, the relationship between prescriptivity and sonic indeterminacy must be specified more precisely. Not all forms of action-based notation lead to unstable or unpredictable results. In many cases, clearly defined actions produce relatively stable and reproducible sonic outcomes, particularly when the causal relationship between gesture and sound remains linear and well-controlled. The situation addressed in this paper concerns a different regime of instrumentality, in which increased prescriptive precision does not stabilize the sonic result, but rather intensifies its ontological instability. The unpredictability that emerges is therefore not a matter of interpretative freedom, but produced by the interaction between precisely prescribed actions and the physical limits of the instrument.

The following section develops this argument by examining prescriptive notation more closely – not as an experimental deviation from tradition, but as an extension of practices that have long existed within classical notation. By tracing the prescriptive layers embedded in descriptive systems, it becomes possible to rethink notation as an interface that operates at the level of action rather than representation.

4. PRESCRIPTIVE LAYERS WITHIN CLASSICAL NOTATION

While musical notation is often characterized as a descriptive system that specifies sonic parameters, a closer examination of classical notation reveals a dense network of prescriptive elements embedded within it. These elements do not describe sound as such, but rather instruct performers how to act upon their instruments. They operate at the level of gesture, technique, and bodily engagement, and their sonic consequences remain only partially predictable.

From this perspective, classical notation already functions as a hybrid system in which descriptive and prescriptive dimensions coexist. Pitch, duration, and dynamics may form the structural backbone of the notational system, but they are continually supplemented by instructions that prescribe instrumental actions without fully determining their sonic outcome. These prescriptive layers are often treated as secondary or auxiliary, yet they play a decisive role in shaping musical sound.

One of the most evident examples can be found in string techniques such as *sul ponticello*, *sul tasto*, *col legno*, or

flautando. These indications do not describe a specific timbre in acoustic terms; rather, they prescribe a spatial relationship between bow and string. The resulting sound – richer in overtones, noisier, thinner, or more unstable – emerges from the action itself and varies depending on the instrument, performer, and context. The notation specifies an action, not a sound object.

Articulation marks such as *staccato*, *legato*, *spiccato*, *tenuto*, or *marcato* further illustrate this prescriptive dimension. These symbols do not correspond to fixed acoustic profiles but to families of gestures that shape the temporal and energetic character of sound production. Their interpretation relies on embodied performance traditions rather than on precise sonic definitions. As such, articulation marks operate as interfaces between notation and bodily technique.

Ornamentation provides an especially revealing historical example. Baroque ornaments – trills, mordents, turns, *appoggiaturas* – were initially not fully specified in notation but relied on shared performance conventions and improvisatory elaboration [9]. Even when standardized symbols were introduced, they functioned less as descriptions of exact pitch sequences than as prompts for gestural procedures. Ornamentation thus represents an early form of prescriptive notation that mediates between structure and improvisation.

Breath marks, bowings, fingerings, and pedaling indications extend this logic further. These markings explicitly address the performer's body and instrumental interface rather than the sonic surface of the music. Fingerings, in particular, reveal the instrument-specific nature of prescriptive notation: the same pitch sequence can be realized through multiple physical strategies, each producing subtle differences in timbre, articulation, and resistance. By prescribing a fingering, the notation intervenes directly in the performer's bodily technique.

Crucially, these prescriptive elements are rarely perceived as undermining the determinacy of classical notation. On the contrary, they are often understood as contributing to precision and control. Yet this precision operates at the level of action rather than sound. The more precisely an action is prescribed, the more clearly the composer articulates a field of performative causality – while simultaneously acknowledging that the sonic result cannot be entirely fixed in advance.

This observation allows for a redefinition of prescriptive notation that departs from its common association with experimental or graphic practices. Prescriptivity does not imply vagueness or indeterminacy *per se*. Instead, it designates a mode of notational thinking in which the primary object of inscription is instrumental action. Sound, in this framework, appears as an emergent phenomenon rather than a pre-specified target.

Seen in this light, classical notation already contains the conceptual tools required to engage with unstable, noise-based, or process-oriented sounds. What changes in contemporary practice is not the presence of prescriptivity, but its degree and emphasis. When sound production becomes more unstable or when instrumental techniques operate at the limits of control, the prescriptive dimension of notation moves from the background to the foreground.

This shift sets the stage for the use of minimally extended notational strategies that intensify prescriptive instruction without abandoning the logic of classical notation. Rather than replacing descriptive notation with alternative systems, such approaches work from within the tradition, exploiting its latent prescriptive layers. The following section examines how this strategy can be conceptualized philosophically, and how prescriptive notation can be understood as operating within a process-oriented, immanent ontology of music.

5. PRESCRIPTIVITY, ACTION, AND VIRTUALITY

The prescriptive dimension of musical notation acquires its full conceptual significance when considered within a process-oriented ontology of music. If sound is understood not as a fixed object but as an emergent event, then the function of notation cannot be limited to the representation of sonic outcomes. Instead, notation operates at the level of action, configuring the conditions under which sound may occur. This shift can be productively articulated through Gilles Deleuze's philosophical distinction between the virtual and the actual.

In Deleuzian terms, the virtual is not an abstract or unreal domain, but a structured field of potentialities that are real insofar as they condition what can become actual [5]. The actual, by contrast, is the concrete manifestation of these potentials in a specific situation. Crucially, the relationship between the virtual and the actual is not one of execution or resemblance: the actual does not represent the virtual, but actualizes it through processes of differentiation.

Applied to musical notation, this distinction allows for a rethinking of prescriptivity beyond the opposition between determinacy and indeterminacy. Prescriptive notation, as understood in this paper, does not specify actual sounds; it articulates a virtual field of instrumental actions. These actions – fingerings, contact points, pressure, airflow, resistance – constitute real tendencies that shape what kinds of sounds may emerge, without fixing their precise sonic identity in advance.

This perspective clarifies an apparent paradox that has been identified in existing accounts of prescriptive notation. As Mieko Kanno has convincingly argued, no musical notation can be purely prescriptive or purely descriptive, and the balance between these dimensions is historically variable [10]. One of the most significant contributions of her work is the observation that the more prescriptive notation becomes, the more interpretative labor it demands from the performer. In this view, notation is not a transparent set of instructions, but a site of negotiation between compositional intention, instrumental affordances, and performative realization.

This paper fully endorses this diagnosis. However, rather than treating this paradox primarily as a pragmatic or epistemological tension within notational practice, it proposes an ontological rearticulation of the problem. The increased interpretative demand associated with prescriptive notation is not accidental, nor does it result from insufficient specification. It is a structural consequence of notation operating at the level of action rather than sound.

From a Deleuzian perspective, prescriptive notation structures a virtual field of instrumental actions whose actualization cannot be predicted in advance. The more precisely these actions are prescribed, the more clearly the virtual field is articulated – yet without collapsing it into a single actualization. Interpretative work, in this context, does not consist in choosing between predefined sonic outcomes, but in actualizing action-based potentials whose sonic consequences remain immanent to the performance situation.

In this sense, the paradox identified by Kanno – the more prescriptive the notation becomes, the more interpretative work it demands – is not a contradiction, but a necessary consequence of action-based notation engaging directly with the virtual. Precision at the level of action does not reduce openness at the level of sound; on the contrary, it intensifies processes of differentiation. The notation prescribes what to do, not what will be heard.

This ontological framing also allows for a refinement of the notion of notation as a site of negotiation. While negotiation can be understood pragmatically as a dialogical process between composer and performer, it can also be conceived more fundamentally as an immanent process in which sound emerges through the differential actualization of prescribed actions. In this latter sense, negotiation does not take place about sound, but within the process of its becoming.

Prescriptive notation, therefore, does not negotiate sonic meaning in advance, nor does it aim to stabilize musical identity. Instead, it activates conditions under which sound can arise as an event. The openness associated with such notation is not the result of ambiguity, under-specification, or reduced precision, but of the impossibility of fully predicting the sonic consequences of precisely prescribed instrumental actions.

This framework also allows for a re-evaluation of the relationship between notation and improvisation. Improvisation, rather than standing in opposition to notation, can be understood as operating within a virtual field structured by technical knowledge, embodied habits, and situational constraints. Prescriptive notation externalizes and formalizes aspects of this field, translating tacit instrumental knowledge into explicit performative instructions. In doing so, it establishes a continuity between composed and improvised practices without reducing either to the other.

Seen through this lens, prescriptive notation should not be positioned at one extreme of a spectrum ranging from control to indeterminacy. Rather, it constitutes a distinct mode of notational thinking in which determinacy is relocated from sound to action. This relocation explains why prescriptive notation can be simultaneously rigorous and open, demanding and unstable, precise and unpredictable.

This conceptualization provides the theoretical foundation for the compositional strategies discussed in the following section. By employing minimally extended classical notation to prescribe instrumental actions while allowing sonic outcomes to remain immanent to performance, the works examined here enact prescriptivity not as negotiation over representation, but as a catalyst for musical becoming.

6. PRESCRIPTIVE NOTATION IN COMPOSITIONAL PRACTICE: CASE STUDIES

The prescriptive approach to notation outlined in the preceding sections is grounded in the author's compositional work for trumpet, where classical notation is employed as a primary framework while being minimally extended to articulate action-based instrumental processes. Rather than introducing alternative or graphic notational systems, this practice intensifies prescriptive elements already latent within the traditional five-line staff, allowing instrumental actions to be specified with high precision while leaving their sonic outcomes inherently open.

Across the following examples, notation consistently operates at the level of instrumental action rather than sonic description. Pitch, rhythm, and articulation retain their conventional functions, but they are strategically reoriented to serve as reference points for the coordination of bodily gestures, airflow, valve technique, and spatial manipulation of the instrument. Sound emerges as an immanent consequence of these actions rather than as a pre-specified object.

6.1. Example 1: Bell Manipulation as a Prescriptive Continuum

In several works, manipulation of the trumpet bell – either by hand or using a plunger mute – is notated using simple prescriptive symbols placed above the staff. The symbols “o” and “+” indicate a fully open and fully closed bell respectively, while arrows between them prescribe gradual transitions between these states. Intermediate degrees of openness are indicated by partially shaded symbols (see Figure 1).

Crucially, these symbols do not describe timbre or dynamic change; they prescribe spatial and physical actions. The resulting colorations of sound depend on numerous contingent factors, including air pressure, embouchure stability, and acoustic feedback. In extended tones, the same symbols are sometimes rhythmically articulated using note stems without heads, allowing the opening and closing of the bell to be precisely coordinated in time independently of pitch specification.



Figure 1. Prescriptive notation of gradual bell opening and closure on trumpet.

Here, notation functions as a temporal organization of continuous action. The sound produced cannot be fixed in advance, yet the performer's actions are highly constrained and structured. This example demonstrates how prescriptive notation can articulate a process of becoming rather than a static sonic state

6.2. Example 2: Action Layering Through Slide, Trill, and Approximate Pitch

A second prescriptive strategy involves rapid directional motion within the harmonic series, produced through controlled overblowing while simultaneously executing valve trills. In these passages, notation specifies a clearly defined starting pitch, followed by a slide line leading toward an arrow-shaped notehead indicating an indeterminate extreme register rather than a fixed target pitch. The entire gesture is embedded within a strictly defined rhythmic duration and marked with a trill indication above the staff (see Figure 2).

This notational configuration coordinates multiple layers of action: directional airflow modulation through the harmonic spectrum, rapid valve alternation, and temporal precision. The slide does not represent a glissando in the conventional sense; it prescribes a process of overblowing from one register toward another while the trilling action destabilizes pitch identity. The approximate pitch notation explicitly suspends descriptive accuracy at the level of sound while maintaining prescriptive clarity at the level of action.

Interpretation in this context does not involve expressive shaping of a stable sound, but continuous regulation of unstable instrumental behavior. The performer must negotiate resistance, feedback, and threshold states in real time, actualizing a virtual field of action structured by the notation.

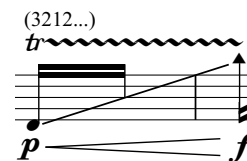


Figure 2. Combined prescriptive notation of slide motion and trilled articulation within a fixed duration, emphasizing action-based control over approximate pitch outcome.

6.3. Example 3: Hyperkinetic Prescriptivity and Sonic Instability

The most extreme application of this approach occurs in a section characterized by sustained hyperkinetic motion. Here, the trumpet part unfolds in continuous rapid activity, notated within the traditional staff using dense rhythmic groupings such as sextuplets and quintuplets at fast tempi. Only selected pitches – marked with noteheads in parentheses – appear as points of directional inflection, indicating moments where the motion reverses from ascending to descending within the harmonic space, or vice versa. Between these points, slide lines indicate continuous motion through the overtone spectrum.

Simultaneously, valve actions are precisely specified using numerical valve indications aligned with the rhythmic structure. As a result, the performer executes two tightly coordinated layers of action: rhythmically articulated valve movements and continuous pitch-space traversal through overblowing (see Figure 3). Despite this high degree of prescriptive precision, the resulting sound is highly unstable, noisy, and resistant to identification in terms of traditional pitch-based musical categories.

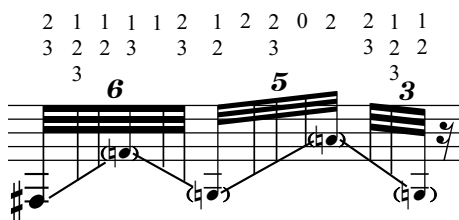


Figure 3. Highly prescriptive rhythmic notation coordinating valve actions with continuous pitch movement across the harmonic spectrum, producing unstable sonic results.

This section exemplifies the central claim of the present study. Maximal prescriptive determination at the level of action produces minimal stability at the level of sonic identity. The notation functions as a machine for generating events rather than reproducing sounds. Importantly, the precise rhythmic articulation also enables close coordination with other instrumental parts, which engage in similarly rapid, complementary motions, establishing a form of contrapuntal interaction based on action rather than pitch.

6.4. Prescriptive Precision and Immanent Openness

Across these examples, prescriptive notation does not reduce interpretative freedom but relocates it. The performer's task is not to realize a predefined sonic object, but to sustain a dynamic relationship with unstable sound-producing processes. Interpretation becomes a matter of managing thresholds, resisting collapse, and navigating instrumental limits.

By maintaining the logic of classical notation while selectively extending its prescriptive capacity, these works demonstrate that action-based notation can operate with high precision without sacrificing openness. Sound remains immanent to performance, emerging through the actualization of virtual action-fields structured by the score.

This compositional practice thus embodies the ontological framework developed earlier in the paper. Prescriptive notation is neither a deviation from tradition nor a strategy of indeterminacy, but a means of articulating music as an event whose identity resides in processes of action and becoming rather than in fixed sonic forms.

7. CONCLUSION

This paper has proposed a rearticulation of musical notation beyond its conventional representational function, approaching it instead as a prescriptive interface that organizes instrumental action and enables the emergence of sound as an event. By examining the prescriptive layers embedded within classical notation and situating them within a performative and process-oriented ontology of music, the study has argued that precision in notation need not imply determinacy at the level of sonic outcome.

Drawing on improvisation, historical practices, and contemporary philosophical frameworks – particularly Deleuze's concept of virtuality – the paper has shown that prescriptive notation operates at the level of action rather than sound. The paradox identified in existing literature,

whereby increased prescriptivity intensifies interpretative labor, has been reframed not as a contradiction but as a structural consequence of action-based notation engaging with unstable sound-producing processes.

Through case studies drawn from the author's compositional practice for trumpet, the paper has demonstrated how minimally extended classical notation can prescribe highly specific instrumental actions while leaving sonic results immanent to performance. These examples illustrate that prescriptive notation does not negate compositional control, nor does it collapse into indeterminacy; rather, it relocates determinacy from sonic representation to performative causality.

By maintaining continuity with the logic of traditional notation while intensifying its prescriptive dimension, this approach offers an alternative perspective on the relationship between composition, improvisation, and performance. It suggests that notation can function not as a blueprint for musical objects, but as a catalyst for musical becoming – one that activates virtual fields of action whose actualizations remain contingent, situated, and irreducible to fixed sonic identities.

8. REFERENCES

- [1] L. Goehr, *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Oxford, U.K.: Oxford Univ. Press, 1992.
- [2] C. Small, *Musicking: The Meanings of Performing and Listening*. Hanover, NH, USA: Univ. Press New England, 1998.
- [3] B. E. Benson, *The Improvisation of Musical Dialogue: A Phenomenology of Music*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [4] D. Borgo, *Sync or Swarm: Improvising Music in a Complex Age*. New York, NY, USA: Continuum, 2005.
- [5] G. Deleuze, *Difference and Repetition*, P. Patton, Transl. New York, NY, USA: Columbia Univ. Press, 1994.
- [6] N. Cook, *Beyond the Score: Music as Performance*. Oxford, U.K.: Oxford Univ. Press, 2013.
- [7] C. Cox, *Sound Art and the Sonic Unconscious*. Cambridge, MA, USA: MIT Press, 2014.
- [8] B. Ferneyhough, "Shaping sound," in *Sound*, P. Kruth and H. Stobart, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2000, pp. 114–20.
- [9] F. Neumann, *Ornamentation in Baroque and Post-Baroque Music: With Special Emphasis on J.S. Bach*. Princeton, NJ, USA: Princeton Univ. Press, 1978.
- [10] M. Kanno, "Prescriptive notation: Limits and challenges," *Contemp. Music Rev.*, vol. 26, no. 2, pp. 231–254, 2007.

PLAYING THE ACTION, HEARING THE RESULT: PRESCRIPTIVE NOTATION AND BASSOON TECHNIQUE IN AARON CASSIDY'S *25 APRIL '05*

James Aylward

Edith Cowan University
jaylward@our.ecu.edu.au

ABSTRACT

The expansion of extended instrumental techniques has increasingly exposed the limitations of traditional Western musical notation, particularly its prioritization of pitch and rhythm over timbre, physical action, and performance process. This paper examines these limitations through a case study of Aaron Cassidy's bassoon solo *25 April '05* (2025), focusing on the composer's use of a highly prescriptive tablature-based notational system. Drawing on Charles Seeger's distinction between descriptive and prescriptive notation, as well as subsequent discourse on mixed notational systems, the article situates Cassidy's work within broader debates surrounding the communication of instrumental technique. Cassidy's score separates the activities of the breath apparatus and the fingers across two staves, employing graphical notation, colour-coding, and tablature to prioritize physical and choreographic action over fixed sonic outcomes. Through detailed analysis of this system, the paper argues that Cassidy's notation creates a sophisticated method of communication that actively engages the performer while creating conditions for instability and variability. The notation stimulates the exploration of transitive and unpredictable sounds intrinsic to the bassoon. While acknowledging the practical challenges posed by such a system, the article suggests that *25 April '05* offers a compelling model for rethinking notation as a generative framework, expanding contemporary approaches to instrumental technique and composer-performer interaction.

1. INTRODUCTION

The expansion of instrumental technique is increasingly testing the bounds of traditional notation. As composers delve more into extended techniques, new notational approaches are required to integrate these practices into the musical language. Traditional notation, which prioritizes pitch, harmony, and metric rhythm, presents music in a form that a trained musician can readily imagine. However, shifts in compositional thinking that foreground timbre, physical action, or alternative approaches to the

instrument remain only partially accommodated in this system. This structural limitation may hinder the development of extended techniques, which are often forced into a framework that is not totally compatible or idiomatic.

This paper argues that Aaron Cassidy's tablature innovations in his bassoon piece *25 April '05* (2025) [1] reconfigure instrumental writing by prioritizing physical action over sonic outcome. Cassidy moves beyond tablature as merely a system of technical instruction, developing instead a more complex and expressive formal language. In contrast to the use of tablature in works by, for example, Billone and McCormack, his notation establishes a richer mode of musical communication.

Historically, tablature has often been regarded as a less sophisticated means of musical communication, associated with pre-literate traditions and systems that primarily indicate the physical actions required to produce sound. However, in the twentieth and twenty-first centuries, composers have revisited tablature for its capacity to engage directly with the instrument as a physical interface. Cassidy creates a sophisticated grammar and syntax that add layers of meaning and encourage a mode of performer engagement distinct from other notational systems. This approach aligns with his prioritization of physical and choreographic elements over purely sonic outcomes. As Mieko Kanno suggests, an openness in the notational specification of the resulting sound can create greater dynamism and spontaneity, when a precise sonic goal is no longer primary [2]. While tablature can suggest openness in sonic outcome through omission, this is often illusory: highly specific physical instructions may in practice constrain the range of possible results. Cassidy's notation focuses on the performer's actions, placing greater emphasis on embouchure and breath than on finger placement as a means of producing specific sounds. While this approach is not without its limitations, it opens new possibilities for rethinking instrumental technique and its modes of communication.

2. NOTATION

Notation has changed incrementally over hundreds of years, adding layers, subtly reinterpreting the meaning of the symbols and generally increasing precision. However, it still essentially communicates in the same way, foregrounding pitch and time on different axes. New forms of notation are difficult to introduce to musicians trained for many years in traditional systems, and the investment

Copyright: © 2026 James Aylward. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

required to learn an entirely new notational language for a single work often outweighs its perceived benefits. This learning threshold plays a significant role in limiting notational innovation, encouraging composers either to adapt extended techniques to existing systems or to minimize their notational demands. New forms of notation, like different languages, allow the composer to communicate in subtly different ways, opening new perspectives in musical discourse. Innovating notation requires an understanding of its limitations, a point articulated by Charles Seeger through his three ‘hazards’ of Western musical notation (1958):

1. The full auditory parameters of music cannot be fully represented in notated form.
2. Musical writing lags behind developments of speech writing.
3. The failure to clearly distinguish between prescriptive and descriptive writing [3].

Seeger’s first point acknowledges the incompleteness of musical notation. There is only a finite amount of information that can be conveyed with a limited set of symbols. Traditionally, this incompleteness was not considered problematic, as notation functioned primarily as a guide for performers rather than a comprehensive prescription. Over time, however, notation has tended toward increasing levels of instruction and specificity. Even so, there are elements that remain unnotatable, and how the performer navigates these elements can elevate the musician past just being a technician.

This can be observed, for instance, in comparisons of different recordings of works by composers such as Stockhausen, who is often cited as an extreme example of highly detailed notation. The metrics by which a performance is judged are not only determined by absolute fidelity to the notation and an awareness of stylistic traditions but also through intangibles such as energy and mood.

Seeger’s second point expands on the incompleteness and delves into the difficulties of music to convey more than just the ‘notes.’ For example, Carlos Chavez, in his *Toward a New Music* (1937) discusses the inability of ‘contemporary notation’ to notate the subtle intonational inflections or rhythmic complexities of folk music. “The constant small irregularities in time and tempo in folk music cannot be captured in notation with complete fidelity [4].” Articulation, phrasing, timbre and sound morphology are also not well represented and interpretation of these aspects of the music making process relies on a stylistic knowledge of the performer to understand unwritten conventions and tradition.

Acknowledging these points, a composer is confronted with decisions about which parameters to specify and which to leave open. Traditional notation has fostered a bias that favours pitch and duration as the most important parameters in musical discourse. Taken to an extreme, Nelson Goodman, in *The Language of Art*, even describes tempo, phrasing and dynamics as “auxiliary directions whose observance or nonobservance affects the quality of a performance but not the identity of the work,” while also stating that the most brilliant performance with a single wrong note does not count as a compliant instance of the

work [5]. While this stance is difficult to sustain in much contemporary practice, the bias is embedded in the notational system. Aaron Cassidy has described this bias being counterintuitive to his practice and how he has negotiated this problem. In his reworking of *the green is where* (2002), Aaron Cassidy replaces traditional notation with tablature. The specific pitches from the original version were a by-product of a composed physical action; the tablature therefore reflects his compositional intentions more directly than a conventional transcription of a single iteration of his process [6]. Rethinking notation and how music is communicated allows the foregrounding of parameters such as gesture, timbre, and physical action within the performance experience.

Seeger’s third point delves into the mixed methods of communication in musical notation. A descriptive notation attempts to represent the sonic outcome of a passage while a prescriptive notation—often in the form of a tablature—instructs the performer in the physical actions required to produce sound without directly specifying the final result. This dichotomy is, however, more complicated. Traditional Western notation which is often understood today in descriptive terms was actually described in Seeger’s article as prescriptive. Traditional Western notation communicates on different levels with Kanno describing how the system is a mixed system of communication in which different types of instruction operate concurrently [2].

Seeing a note on the page, in principle, does not share any information about the production of said note. But instructions such as *pizzicato*, *sul ponticello* or harmonics or even slurring, while conjuring a different image of the sound are directly related to the action of producing the sound.

However, there are also implicit prescriptive instructions that a performer follows when reading traditional Western notation. As a bassoonist if I see a notated C3, I think to close the first 3 holes and blow at a certain pressure to create this sound. The octave higher is an instruction to cover the same 3 holes but to apply more air pressure to filter out the lower partials. Of course, I could also find an alternative fingering combination that would produce the same pitch, but this choice is mostly initiated by the performer to better realize a sound ideal for a particular passage and outside the thinking of many composers. The use of extended techniques has, in particular, increased the need for prescriptive instructions. Extended techniques expand the thinking beyond the ideas of pitch, harmony and rhythm and are not part of every musician’s vocabulary so have to be communicated in a prescriptive way. Secondly, extended techniques are usually derived from an analysis of the performance technique and demonstrate an alternative manner of playing being aided by a more prescriptive instruction. Adding too much information to a score can reduce readability especially as instructions for fingerings (for example) do not mix always easily with the language in use. As music incorporating extended techniques become more ingrained into the repertoire these symbols can take on a more descriptive quality.

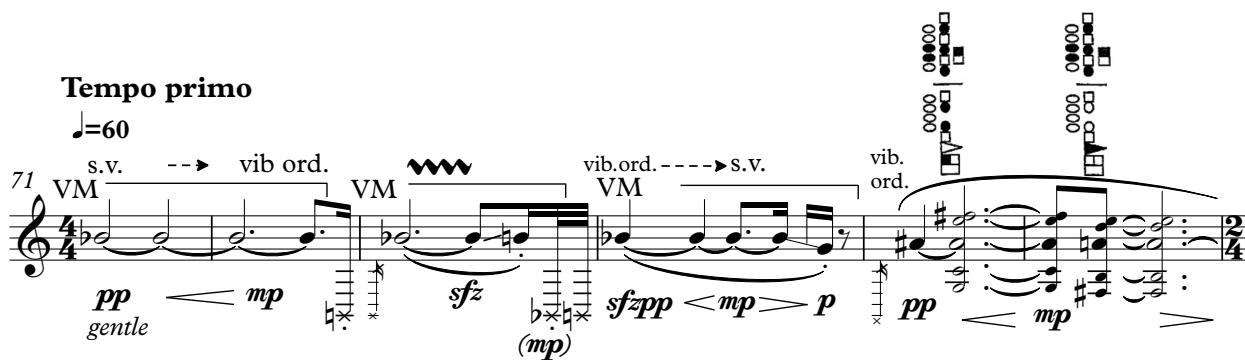


Figure 1. Djordjević – *Nailing Clouds* (2019). Bars 71-76 [8]

The coexistence of descriptive and prescriptive elements does not, however, have to be seen as a failure as Seeger suggests. Rather, it introduces a productive complexity into the score, and an awareness of these different modes of instruction can enable performers to develop more nuanced and sophisticated interpretations.

3. NOTATION APPLIED TO THE BASSOON

Notation and Composition determine each other. Differentiate between creating a language in order to say something and evolving a language in which you can say anything. A musical notation is a language which determines what you can say, what you want to say determines your language. (Cardew) [7]

The testing of the limitations of traditional notation accommodating extended techniques can be seen in Milica Djordjević's *Nailing Clouds* (2019) [8] (Figure 1). Djordjević draws on a wide range of techniques described by Gallois in *The Techniques of Bassoon Playing* (2009) to create a rich palette of timbral possibilities [9]. The example shows a passage that is particularly challenging. There are many instructions which apply to the quality of sound such as the *Velvet Sound* (VM) marking as well as different types of vibrato, mood and dynamic instructions. The note B flat has an obvious structural function (as also implied by traditional notational practice) being repeated with different kinds of vibrato, with velvet sound and then normally before breaking into a multiphonic in bar 75. The velvet sound is described in Gallois' book as resembling "brass instruments played with velvet mutes [9]." He also describes a technique of blowing that can produce that sound but concedes the intonation will be *around a quartertone higher* [9]. The important structural element of the note would suggest to me that the intonation is important as well as finding an effective timbre for the

velvet sound. My approach has been to see the passage in descriptive terms and find solutions through alternative fingerings that would fulfill these aims (and also to be able to glissando in bar 74). One can see that the traditional interpretations of the established symbols are becoming more flexible, creating choices for the performer which may or may not reflect the intentions of the composer. Through working with Djordjević I have also learnt that her use of accents should be as aggressive and hard as possible—far above what the traditional way of playing requires. While she uses the traditional accents and sforzando instructions often supported with extra textual instructions such as, "dark, rough / unrefined timbre [8]," the extremity of what she wants would be difficult to know from the notation alone. A knowledge of the implicit stylistic considerations is a part of all readings of notation but for contemporary pieces this cannot be assumed as standard knowledge and practice.

As Cardew wrote, notation and composition determine each other. Djordjević, by working with traditional notation retains many aspects of a traditional approach to music albeit expanded timbrally with extended techniques. Her use of multiphonics is influenced by traditional notational thinking and as a result they are not fully exploited. They are often used as a distortion effect transitioning from single pitches or they are used as rest points to contrast the melodic material. The notation with the fingerings above the notes imbues a perceived awkwardness that is better suited to slower treatments. The piece ends in a chorale like passage of multiphonics which treats each multiphonic as a static block of sound. While this passage is particularly beautiful, the representation of these sonorities as chords does not accurately reflect their perceptual reality, which is often closer to noise than to a harmonically defined structure.

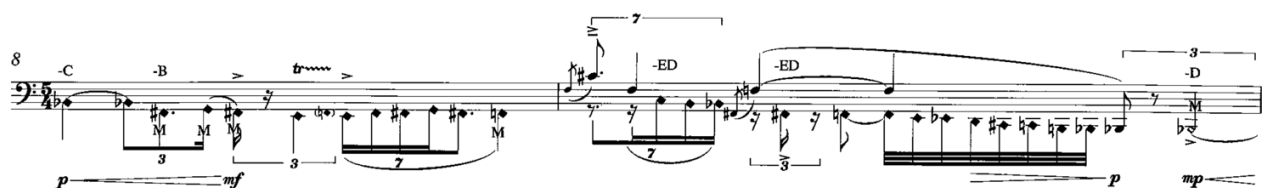


Figure 2. Lim, *Axis Mundi* (2013). Bars 8-9 [10]

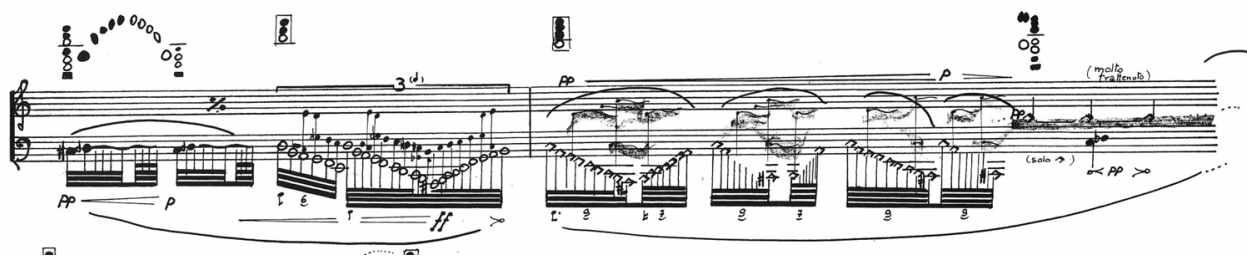


Figure 3. Billone, *Manda* (2003) From p. 5 [11]

There is no ideal notation that can be applied to all kinds of music, but notation can influence how a performer approaches a piece and can foreground different aspects of performance. Liza Lim in *Axis Mundi* (2013) [10], combines a form of tablature with normal melodic notation. This suits the narrative of the piece which is meant to alternate between the real world and spiritual realm. The instructions are given as altered fingerings—a base alteration to remove or add a finger while the music goes further (notated with diamond noteheads). (Figure 2). The alterations are simple but are enough to distort the sound in various ways. Due to its size, the bassoon has smaller holes than on the other woodwind instruments. Opening holes higher on the tube than the designated note thus does not totally change the resonating column leading to a combination of register shifts, multiphonics as well as bisbigliando effects. Seeing a representation of fingerings as noteheads is more intuitive to read and Lim can use the novel sounds in a more dynamic way similar to her treatment of the more standard material. The tablature, like Kanno suggests removes the pressure of finding specific sounds so the performer's attention can be more directed to the musical gestures.

A similar technical approach but with a very different result can be seen in *Manda* (2003) by Pierluigi Billone [11]. The use of handwritten tablature alongside the resulting sound produces a markedly different impression for the performer signalling an absence of traditional melodic thinking in favour of an alternative mode of musical discourse. The dual presentation of prescription (fingering/action) and description (resulting sound) asserts a strong authorial presence, consistent with Billone's well-known, in-depth investigation of instrumental practice. (Figure 3).

Importantly, Billone's tablature demonstrates how the material is generated through simple physical manipulations of the resonating column. It is economical, highlighting only the changing elements of the fingering and providing occasional reminders. Traditional fingering patterns are also notated with traditional 'pitches.' The

descriptive staff complements this by capturing the essential character of each sonority with clarity and precision, leaving little doubt as to Billone's intentions. Together, the two notational layers guide the performer through both the generation of material and its execution, emphasizing the interplay between technique and sonic outcome.

A more extreme attempt to innovate notation can be seen in Timothy McCormack's *Body Matter* (2015) [12] (Figure 4). McCormack eliminates the fingering pictograms that are utilized in many pieces using extended techniques. McCormack, instead, assigns each key and hole to a 'pitch' on the staff. A fingering combination is presented firstly as a red notehead which shows a basic fingering with open noteheads prescribing the removal of a key and plus noteheads showing the adding of a key. A green notehead shows the dominant pitch in a sound also giving an idea of how to blow into the instrument. An extra beam above the staff shows "embouchure pressure [12]" and reed position. This instruction is largely redundant, since the other instructions already convey implicitly much of the same information. But it also serves to emphasize the continuous, sustained nature of the piece. McCormack's notation may seem convoluted, but it perfectly encapsulates the different sound world he is inhabiting. While Billone's treatment of sound can be seen as very natural—there is a living, vocal quality to his music, McCormack is looking for a more otherworldly character that is reflected in his notation. McCormack's notation also allows a more irregular treatment of the material, and he utilizes many unusual and awkward fingering combinations. But ultimately the focus, like Billone, is on the production of specific sounds thus fulfilling a similar role to most other notations.

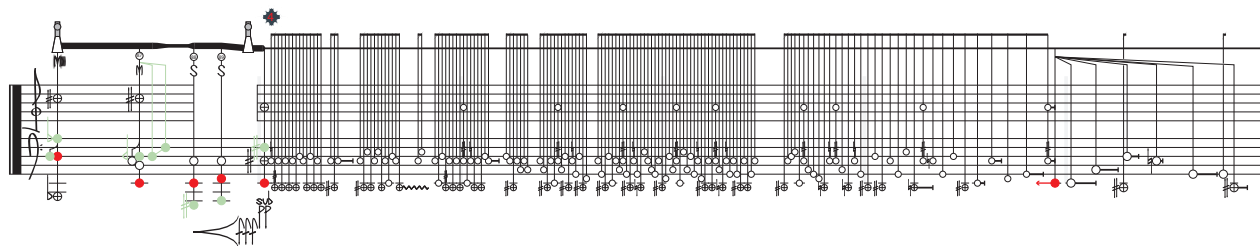


Figure 4. McCormack, *Body Matter*. (2015) from p.6 [12]

4. CASE STUDY: 25 APRIL '05

Aaron Cassidy's *25 April '05* illustrates how a composer can develop a notational system that eschews the traditional priorities and foregrounds the physical action. Cassidy's use of tablature is a deliberate way to engage the physicality of the bassoon but is not just a system to find unconventional sounds. He creates a system that communicates on multiple levels, similar to the mixed methods found in traditional notation. This allows for nuanced interpretation, varied energy states, and the discovery of unique and often unstable sounds. Cassidy has instead created conditions where sound instability happens naturally. The results can be different through different iterations, but he finds sounds which are transitive and unstable by exploring confluences where different parts of the technique interact in unexpected ways creating a live exploration of the instrument.

While Kanno emphasizes a freedom that comes from a focus on the process rather than the end result, the specificity of Cassidy's instructions can be more restricting—through their abundance and difficulty—but this helps create the energy and tension in his music. The preparation process is very different. Where in a traditional score there is an ingrained approach to focus on quality of sound, intonation, and the right notes, Cassidy's approach requires a focus more on action and movement. Shapes of notes, different playing techniques, and timing assume more structural importance as the main discernible instructions. The score is deliberately written to sound 'messy' with techniques that contradict each other and many actions unsynchronized. Instead of a turn of phrase or timing that is left open, the resulting sound is open, with the discipline being directed to the choreography of the performance. Cassidy's approach is particularly well suited to the bassoon whose unpredictability can be notorious for composers looking for specific effects. Liza Lim writes of her reluctance to use catalogued multiphonics calling it an "extremely inexact science [13]." Unstable fingering combinations vary depending on their context as well as from player to player. The smallest change in the embouchure or air pressure can change the sound drastically. Additionally, these fingerings are not well represented by a static translation on paper that neglects their flexibility and instability.

25 April '05 is based on the overpainted photograph by Gerhard Richter (Figure 5) and is a companion piece to Cassidy's solo E flat clarinet piece, *27 June '09* (2021). He builds on a similar tablature slightly adjusted for the differences in the bassoon technique. The score to *25 April '05* is notated on two staves: one for the "activities of the mouth/breath/tongue/embouchure [1]" and the other for the activities of the fingers. This second staff employs two types of notations, one working with tablature and the other with using pitches to indicate fingerings.

The breath staff is notated graphically, avoiding traditional symbols. Notes are represented as shapes, with thickness indicating dynamics and length indicating duration. This form of notation conveys a wealth of information: the shape of a note provides guidance for articulation as well as its internal dynamic contour. It also allows for unconventional shapes, such as blocks or

"reverse" sounds. This is an important consideration for instrumental technique, but is not communicated in detail in traditional scores which require that performers have a knowledge of style and convention. It should be noted that the assumption could be made that the sound would resemble these shapes, but these shapes become manipulated with different fingerings and register changes not always sounding like single blocks of sound they might seem to represent.



Figure 5. Gerhard Richter, *25 April '05* (2005) [14]

Cassidy attempts to combine spatial notation with the metrical beat hierarchy with a line showing the approximate tempo for each beat. Cassidy describes this as aiming "to disentangle rhythm from counting [1]." Cassidy additionally employs a colour system which aids the legibility of the music giving an immediate association to process that would not be so visceral using the one colour. The colours also allow for many different types of instructions to be presented on a single breath staff. (Table 1) There are also other instructions like reed positions and percussive techniques that are shown with different symbols.

Colour	Meaning
Black	Default colour. Interpreted as a direct contrast to the orange. Clean sound
Orange	Distort the sound- try to play multiphonic sounds
Blue	Guttural sound/throat flutter
Purple	Direct instruction to a mouth position
Green	'Downward pitch bends'
Pink	Sucking or kissing noise
Red	Play with teeth on reed
White	With reed off. Blow into crook
Grey	Breathy/Diffuse/unfocused sound

Table 1. Colour Instructions in Aaron Cassidy's *25 APRIL '05* (2025)

What is structurally significant in Cassidy's prescriptive notation for the air apparatus is the use of different paradigms to instruct the performer. Cassidy generally avoids a register prescription focusing on other ways to measure the apparatus of the mouth and air column. Cassidy's orange colour instructs the performer to saturate the sound to create multiphonics. This can be applied differently to different notes, but the general instruction is to try to break the sound. Even with fingering where this will not occur easily, it forces a change in sound to something that is not how the performer would usually play. Conversely, the normal pitch instruction (default black) on an unstable fingering encourages jumps in registers where I have interpreted it to mean the cleanest sound possible.

The green part of the score is a supplemental colour that is added to orange or black or blue instructions. It indicates a dropping of pitch while trying to maintain the sound type (inevitably changing sound type in some sections). Bending down is more natural on the bassoon than bending up so the opposite instruction is given as an orange shade and instructs the performer to try to find higher overtones in the sound preferably leading to distortion. (Hence the use of orange again) Cassidy is trying to play on the instability and unpredictability of the bassoon so the attempt to realize the effects is often more important than the sound itself and the parts of the score where this is more difficult produce interesting results that would not be easily notated in another way.

Cassidy's purple instruction has some similarity to the green but is a more absolute performance instruction, giving no indication at all about the resulting sound. Blue functions in a similar way, specifying a manipulation of the sound without defining its sonic outcome. Unlike in *27 June '09* [15], where black functioned as a default mode of sound production, *25 April '05* introduces orange as an explicit opposite to black. As a result, blue and purple (and red) acquire the additional significance of withdrawing any guidance toward a particular sound type.

Sections governed by blue and purple therefore operate with a different internal dynamic: sound becomes a byproduct of technique, and technique itself becomes the determining factor of interpretation. When placed

alongside the black, orange, and green instructions—whose goals are defined but inherently unstable—this creates a subtle but significant dichotomy of energy state. The performer shifts from attempting to control sound within a given parameter type to working with the instrument's natural tendencies. While interesting sonic results may be emphasised, they remain secondary to the application of the prescribed technique.

The fingering instructions are also presented in two ways. The first is as a tablature giving each finger its own directions and the second as pitched notation where the implicit fingering combinations are assumed. The contrast between these two notations also has a structural importance where different explorations of the instrument are made. As Evan Parker has described about his own practice, "The saxophone could just as well be seen as a closed tube that can be opened in various ways as an open tube that can be closed in various ways [16]."

The tablature, being more unpredictable in its sonic outcome, explores the instrument through different, non-traditional fingering combinations. In these passages, fingerings that correspond to certain pitches lose their significance as functional pitches, exemplifying Kanno's freedom of sonic outcome.

The pitched parts can be muddled by other manipulations of the sound but explore more melodic material which shows that Cassidy has not entirely abandoned the traditional musical parameters. Figure 3 shows Cassidy's use of colours and tablature. The purple technique is introduced in bar 7 after orange is the dominant colour in the first 6 bars. Purple will play a more prominent role in the next few bars.

5. IMPLICATIONS FOR PERFORMER AGENCY

Discussions of notation and its relationship to performer agency often hinge on how the performer's role is framed within the score. Marco Blaauw has described one such framing as that of an executant or "bricklayer [13]." Cassidy's alternative use of notation proposes a different framing, engaging the performer through physical actions that are malleable to the individual body and instrument,

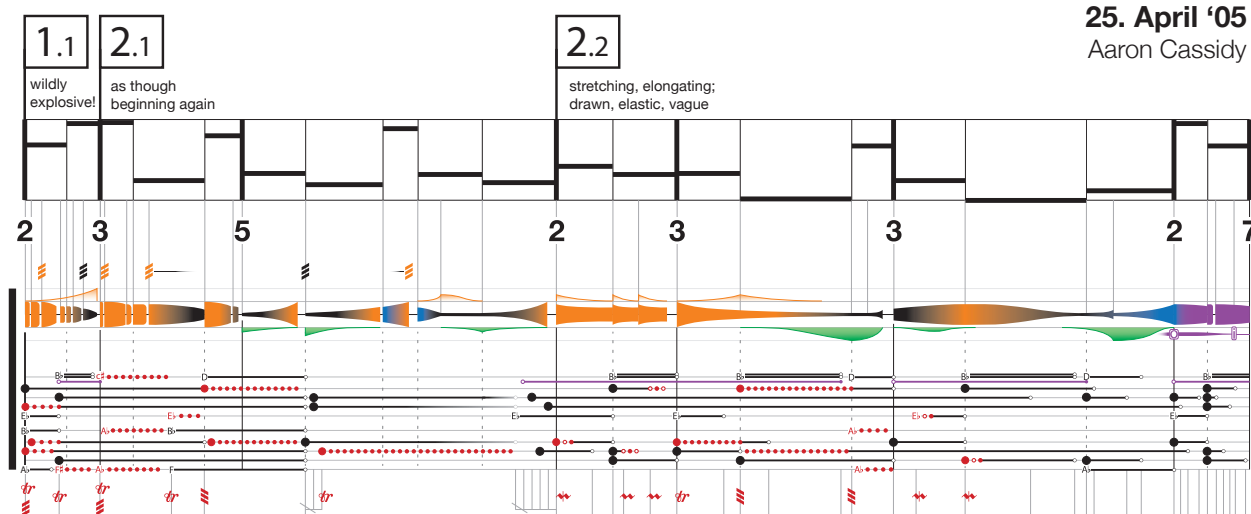


Figure 6. Aaron Cassidy, *25 April '05* (2025) mm 1- 7 [1].

albeit at a highly virtuosic level. He creates a score that fundamentally reprioritizes musical communication. Physical action—particularly the actions of the mouth—assumes a heightened importance, becoming central to the compositional structure. This act of foregrounding requires a loosening of control over other musical parameters that might otherwise distract from Cassidy's primary aims, while simultaneously imparting greater agency to the performer. In this system, technical action assumes structural importance. Motives are presented through repeating fingering combinations even though they may have no predictable direct representation in the sound. The breath staff is also clearly divided into various formal sections using different colours and shapes. Like traditional notation, there are different options for notating different techniques which aid in giving Cassidy's technical instructions a structural function. For example, purple and green can both be used to show the same technical instruction just as a pitch can have its enharmonic equivalents. Cassidy manages to create different layers of the technique which function independently in a sort of virtual polyphony. While other composers have tried the same in traditional notation, the tendency to approach a score as a series of moments can counteract this polyphonic intention. The focus on action encourages the sense of polyphony but as Cassidy says, he is still conscious of the dangers of moment thinking in his highly detailed scores.

Cassidy uses a hybrid rhythmic notation combining a metrical structure with spatial notation, effectively creating a dual rhythmic system. The shapes in the metric structure also assume structural importance, complementing the changing energy states through variations in speed. This relationship can, however, also be obscured when slower beats contain dense activity. While this creates more sophistication in the approach to metric tempo and engages the performer's 'feel' for the rhythm rather than a cerebral approach, the constantly fluctuating beat does not help in creating a sense of flow. The spatial notation also presents choices to the performer in how to interpret timings of certain events. Information requires a certain amount of space on the page to remain practically legible, and this does not always fit proportionally, or neatly into its musical space. However, the inconsistencies increase the scope for interpretation, giving the performer more agency, creating more of a collaboration. At the end of bar 3 in Figure 5, the spatial notation allows for some freedom in rhythmic interpretation with the 'grace notes' not being able to be communicated legibly in their own absolute time. The notation relies largely on graphic representations, but Cassidy still uses open freer prescriptive instructions from conventional notation such as ornamentation. Vibrato and fast tonguing are also represented in a quasi-symbolic/graphic way.

The tablature, while it presents its information in a very clear and structural way, still has certain limitations and probably aptly epitomizes the difficulties of creating a tablature on bassoon. A good technique holistically synchronizes all the different aspects complementing each other to achieve results greater than the sum of the individual parts. For example, changes in air pressure need to be tempered with changes elsewhere to reduce the

negative consequences and this happens in a fluid way. Even the fingerings are adjusted to suit different passages including subtle venting to reach higher registers. This is an essential part of the bassoon technique and something that can be applied to *25 April '05*. Because a tablature does not generally set a sonic goal, the performer could neglect the sound and intentionality in an interpretation. An instruction to change the reed or mouth position comes with an implication of a change of sound quality but this also has to be actively followed by the performer. Just following the instructions will not find all the sonic variations that Cassidy is searching for in his notation.

The complexities of prescribing the embouchure are too much to describe fully on the page but Cassidy resolves this by alternating between different processes of goal-oriented playing and technique-oriented sections. Miniscule changes in this system can have profound impact on the sound so a descriptive element to the instruction is very effective. The keywork is also very complicated on the bassoon and while Cassidy's writing demonstrates some inexperience of the instrument it also helps present his unique perspective not burdened by tradition. Cassidy does an excellent job at providing graphical information, communicating sophisticated instructions on how to play the instrument but presents a dense system that needs much investment to learn. Traditional notation, which is more symbolic, presents information in a more concise and economical way (outside of extended techniques) but has centuries of learned conventions that help it communicate with more immediacy. Dynamics, although simplified and stratified, read easier than an analysis of the sizes of shapes of the musical gestures. And a note on the page conveys an implicit fingering instruction which does not have to be shown constantly. This is demonstrated in the two different processes that Cassidy uses. The tablature is complicated in that it has instructions for the ten digits—also to be combined with the breath system. It is unfamiliar as it deals with atypical fingering combinations, but this allows an exploration into fingering patterns that are outside the realms of traditional patterns, and it is difficult to think of another system that could explore the bassoon in this comprehensive way.

6. CONCLUSION

A tablature system can be difficult to implement on the bassoon and, more generally, risks becoming overly simplistic, neglecting the instrument's complexity. Cassidy, however, has created a system that enables a sophisticated dialogue between composer and performer. Like a spoken language there is also an interesting hierarchy of grammar where for example, the black colour can be manipulated by a reed position manipulation to change the meaning of the original black sound. This also reflects his inspiration of the Gerhard Richter painted over photo where added layers interpenetrate and change the original meanings and implications. His notation reflects his compositional aims, and this phenomenon can also be reflected in other composer's approaches to notation and their musical style. *25 April '05* presents many of the interpretational challenges associated with learning a

complex work, requiring the performer to actively engage with the score and to find solutions to musical problems. This ensures that the piece has significant room for growth over time, and it is hoped that more bassoonists will take up the challenge of exploring it.

Writing for the bassoon can be a daunting proposition, but this piece demonstrates an approach that enables the creation of unique sounds without exhaustive prior study of the instrument. Traditional notation encourages a mode of thinking rooted in nineteenth-century practice, and while composers have challenged many of its assumptions, an inherent bias toward pitch and rhythm remains. New notational approaches allow for alternative ways of thinking that are more closely aligned with contemporary musical concerns. This is arguably particularly relevant for the bassoon; an instrument whose unique design affords an exceptional breadth of timbral, spectral, and gestural possibilities.

Acknowledgments

I would like to thank Aaron Cassidy for composing this piece, for his patience during our collaboration sessions, for his engagement with the idiosyncrasies of the bassoon technique, and for his insightful perspectives on its interpretation. I also thank WAAPA for support in commissioning the piece and particularly Lindsay Vickery for his guidance and assistance during my research.

7. REFERENCES

1. A. Cassidy, *25 April '05*, self-published, 2025.
2. M. Kanno, "Prescriptive notation: Limits and challenges," *Contemp. Music Rev.*, vol. 26, no. 2, pp. 231–254, 2007.
3. C. Seeger, "Prescriptive and descriptive music-writing," *The Musical Quarterly*, vol. 44, no. 2, pp. 184–195, Apr. 1958.
4. C. Chávez, *Toward a New Music*, New York, NY: Dover Publications, 1947.
5. N. Goodman, *Languages of Art: An Approach to a Theory of Symbols*, 2nd ed. Indianapolis, IN: Hackett Publishing, 1976.
6. A. Cassidy, "Determinate Action/Indeterminate Sound: Tablature and Chance in Several recent Works", in *Facets of the Second Modernity*, pp. 17–32, Hofheim: Wolke Verlag, 2008.
7. C. Cardew, "Notation: Interpretation, Etc," *Tempo*, no. 58 pp. 21–33, 1961
8. M Djordjević, *Nailing Clouds*, self-published, 2019.
9. P. Gallois, *The Techniques of Bassoon Playing*, London, UK: Bärenreiter, 2009.
10. L. Lim, *Axis mundi*, Berlin: Riccordi, 2013.
11. P. Billone, *Manda*, self-published, 2003.
12. T McMormack, *Body Matter*, self-published, 2015.
13. L. Lim, "A mycelial model for understanding distributed creativity: collaborative partnership in the making of 'Axis Mundi' (2013) for solo bassoon", in *Performance Studies Network Second International Conference 2013*, Cambridge, UK, 2013.
14. G. Richter, "25 April '05", Gerhard Richter Archive, 2005. [online]. Available <http://www.gerhard-richter.com/en/art/overpainted-photographs/people-74/25-april-05-14295>
15. A. Cassidy, *27 June '09*, self-published, 2021.
16. E. Parker, *De Motu*, 1992. [online]. Available <http://www.efi.group.shef.uk/fulltext/demotu.html>

F-OPERATOR: A PROPOSAL FOR A GENERATIVE NOTATION FOR FREQUENCY FACTORS IN MUSIC

Markus Lepper
semantics gGmbH
Berlin, DE
post@markuslepper.eu

Baltasar Trancón y Widemann
Technische Hochschule
Brandenburg, DE

ABSTRACT

In music theory and historical musicology, tuning systems can be discussed from a variety of angles, ranging from historically given contingent facts or hypotheses, over physiological and psychological experiments, up to integration into an aesthetic theory. All these different approaches define specific collections of objects and calculation rules and live on a symbolic side.

Eventually, these calculations deliver numeric factors to be applied to some reference frequency. These form the factor side. We propose F-operator, a notation system for these values which allows easy identification and labeling of commonly used factors and thus easy comparison of distinct systems from the symbolic side. F-operator is equally well suited for writing, talking, speech recognition, sign language, Braille, and computer representation.

1. PROBLEM ANALYSIS

1.1 Separation of two sides

In music theory and historical musicology, tuning systems can be discussed from a variety of angles, ranging from historically given contingent facts or hypotheses, over physiological and psychological experiments, up to integration into an aesthetic theory. A wide diversity of approaches can also be found in harmonic theories, in physiological research, and in many other realms exploring the perception of music.

All these disciplines can have more or less precise and very different “algebras” (in the widest sense of the word) as their modelling means. The elements of these algebras shall be called *pitches* in the following; all these algebras live on a *symbolic side*.

But eventually all symbolic systems deliver some *factors*. This corresponds to the human way of reception: A particular factor between any two audible frequencies is received as the same interval; a particular sequence of factors between adjacent tones is received as the same melody, inde-

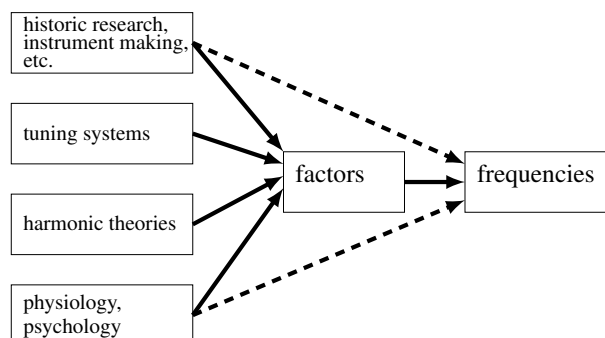


Figure 1. Research realms and their relations to factors and frequencies.

pendent from the absolute starting point. [1, p. 10]¹ The frequencies and factors live on the *factor side*.

See Figure 1 for a simplified depiction of the situation: Most realms indeed talk (mainly) about the factors between frequencies,² only few directly about frequencies: historic research on tuning system does, physiology when addressing duration values of nerve reaction times, etc.^{3 4}

All the more, a concise naming system is helpful. Here we propose the F-operator family, a notation system for the factors on the frequency side, which (a) reflects a minimal and canonical way of their generation, and so (b) supports comparison of and talking about the different systems on the symbolic side.

1.2 The naming problem

The fundamental problem in giving names to frequency factors is that all existing approaches operate on the very

¹ Remarkably, we could not find an earlier scientific publication of this fundamental fact. Even Helmholtz [2] does not mention it; it seems it has been mere folklore ever since.

² In times when the physical concepts of oscillation and frequency were not yet known, these factors were given as ratios of string lengths.

³ And esoterics do: Renold [3] quotes Rudolf Steiner “c[=C3] = 128 Hz = Sonne” [p.108] “[ist] heute der einzige für den westlichen Menschen wirklich geeignete Kammerton.” (“is today for Western people the only appropriate tuning.”) [p. 13].

⁴ In most use cases, also the values on the frequency side exist as *intentions*, as *concepts*. Euler already recognized the twelve-tone equal temperament tuning as an auxiliary means which can be perceived by a listener as representing a just intonation (“Substitutionstheorie” / “theory of substitution” [4, p. 163 ff] [5, p. 125]; see also [6, p. 49]) and contemporary mathematical approaches see those frequencies as the center of fuzzy ranges [7, p. 60].

F 1/2	=	uF 1/2	=	$\sqrt{2}$	=	1.4142...	=	600	cent
		uFn 6	=	$2^{-9} * 3^6$	=	1.4238..	=	611.7300...	cent
		uFn 2 -2	=	$2^2 * 3^2 * 5^{-2}$	=	1.44	=	631.2826...	cent
Fn 0 -1 1	=	uFn 0 -1 1	=	7/5	=	1.4	=	582.5122...	cent
uF 11↑1	=	uFn 0 0 0 1	=	11/8	=	1.375	=	651.3179...	cent

Table 1. Different intervals which all are sometimes called “tritone”.

different symbolic sides to find names for the general and common factor side. Significant examples are:

English wikipedia, lemma “Semitone”, section “Just intonation”⁵: “Just chromatic semitone, chromatic semitone, or smaller [chromatic semitone], or minor chromatic semitone [vs.] Larger chromatic semitone or major chromatic semitone, or larger limma, or major chroma, [vs.] Just diatonic semitone or smaller [diatonic semitone], or minor diatonic semitone, [vs.] Larger diatonic semitone or greater [diatonic semitone] or major diatonic semitone,” — “Descartes used the word schisma [for] the syntonic comma (81:80).”

Ibid., lemma “Diesis”⁶: “As a comma, the above-mentioned 128:125 ratio is also known as the lesser diesis, enharmonic comma, or augmented comma.”

Haluška [7, p. xxix] calls thirteen just fifths modulo octave (in F-operator written as uFn 13 = F-8 13, see below) a “Pythagorean double diminished fifth”, and “512/375” (p. xxvii) (= uFn-1 -3) a “double diminished fifth”. But all these names are meaningful only in some particular pitch calculus on the symbolic side. Figure 2 shows a doubly diminished fifth, which is an appropriate naming from functional harmonic analysis, describing the downward step from D//9- to D//5+, both pitch roles being phenomena with a complicated historical genesis.⁷ This can be realized by a step from cFn -1 -1 to cFn 0 2 of size cFn 1 3.) This is obviously a quite different mental object than the above-mentioned historically neutral frequency factors.

There are prominent examples from the community of micro-tonal composers and researchers:

A text file published by the *Huygens-Fokker Stichting*⁸ lists 604 rational numbers, each with a name but without any explanation, up to the 26-digit

19383245667680019896796723
19342813113834066795298816 .

Another proposal by Keenan (1999)⁹ is much more systematic, but restricted to rational intervals. His table may be of high value for the particular underlying theory, but not in general. A sentence like “I have included the interval 8:13 [...], because it is certainly a better-known representative of the category of *neutral sixth* than is 11:18” shows

⁵ [https://en.wikipedia.org/wiki/Semitone#Just_intonation\[20251103\]](https://en.wikipedia.org/wiki/Semitone#Just_intonation[20251103])

⁶ [https://en.wikipedia.org/wiki/Diesis\[20251103\]](https://en.wikipedia.org/wiki/Diesis[20251103])

⁷ We use the computer-compatible funCode notation for functional analysis, because it is the only with precisely defined mathematical semantics: “D” means the triad of the dominant; “D/” the same without its root pitch class; “D//n” cancels all default triad components and refers to the chord component *n* only. In functional theory, the root of the second chord D/5+79- in Figure 2 is a cancelled b flat. [8]

⁸ [https://www.huygens-fokker.org/docs/intervals.html\[20251222\]](https://www.huygens-fokker.org/docs/intervals.html[20251222])

⁹ [https://www.dkeenan.com/Music/IntervalNaming.html\[20251222\]](https://www.dkeenan.com/Music/IntervalNaming.html[20251222])



Figure 2. Indeed a doubly diminished fifth in Bruckner, First Symphony, Linz version, mv. 1, m. 102.

both: that a certain degree of arbitrariness is involved, and that the naming is tightly related to one particular theory.

1.3 The F-operator approach

The basic idea of F-operator is that a combination of leading alphabetic characters and trailing numbers names the fundamental frequency factors in music theory.

Examples from technical history show that (a) such combinations easily can receive an “auratic identity”, which makes them recognizable and easy to speak and remember, like “Z 80”, “PDP 11”, or “HAL 9000”.

Furthermore, (b) the sequence of numbers shall reflect directly the derivation of the factors from the series of partial tones, an idea proposed (earlier but independently) by Joe Monzo [9, 10], while the leading characters establish the fundamental categories of derivation.

The result is a format neutral to the different symbolic sides and very compact to write and speak. For instance, “F0” is the factor 1.0, normally called “perfect prime”, and “F1” is the “perfect octave”. “uFn 1” is the “just fifth” and “mFn 1” the “just forth”. “uFn12” is the “Pythagorean comma” and “F1/12” the half-tone in 12 EDO tuning.¹⁰

Table 1 shows that F-operator is suited to unambiguously and easily label intervals which are often mixed up.

2. ARCHITECTURE OF F-operator

2.1 Goals and Intentions

The proposal F-operator addresses multiple difficulties in discussing about tuning and harmonic systems, and in many others of the above-listed realms:

[G-1] It allows to use a common notation of the effects on the factor/frequency side, which are independent of all, and common to almost all, generating theories on the symbolic side. This allows (a) a clean separation of these two

¹⁰ 12 EDO means *Equal Division of the Octave by twelve*. We prefer this nomenclature to “12 TET”, because “12 tone equal temperament” implies a particular application context, namely to “temper out” some irregularities from some “original” tuning.

sides, and (b) gives a tertium comparationis for the different generating systems on the symbolic side.

[G-2] Most systems on the symbolic side assign own names to their results on the factor side. This contributes to the overwhelming plethora of synonyms found in literature which shall be replaced, see the examples above.

[G-3] The structure of the numeric label shall reflect one possible method for *deriving* the interval—it shall be *generative*. This provides a more characteristic mental model.

[G-4] There is danger of confusion, especially when discussing the relation between the harmonic spectrum and harmonic theories, between the physical intervals relative to the base frequency, or relative to the next-lower octave of the base frequency, or about all frequency factors which only differ by octaves, etc. For all these cases the *multiple* functions defined by F-operator shall improve clarity.

[G-5] All defined functions are well-formed mathematical objects and can immediately be used for further mathematical constructions, see Table 5 for examples.

[G-6] All expressions shall be written, read, and spoken as easy as possible, with minimal formal noise. They are equally well-suited for Braille, sign language, voice recognition, and computer representation.

2.2 Construction

Table 2 shows the complete construction of F-operator. The type $\hat{\mathbf{F}}$ is its central data structure. Each instance contains *tuples of exponents*, called *f-tuple*, for “frequency tuple” or “factor tuple”. The first line of the table defines $\hat{\mathbf{F}}$ as all total functions from $\mathbb{N}$ into rationals which have only finite support (positions with a value different from zero).

The constructor F takes a finite, possibly empty list of rationals and delivers an instance of $\hat{\mathbf{F}}$: All positions are set to zero and then overridden by the explicit input.

For lean definitions, the technical types of all structures have been chosen as relations, even when they are functional. Function f evaluates tuples from $\hat{\mathbf{F}}$ into real numbers: The contained rational numbers are used as exponents for the sequence of prime numbers $\mathcal{P}$ and the results are multiplied. The assignments between bases and exponents are by their respective position in both sequences: the leftmost exponent in $e : \hat{\mathbf{F}}$ is taken as the exponent of base two, the next that of base three, of five, of seven, etc. The operators $*$, $/$, and $\uparrow$ are implemented pointwise.

All this produces a somehow peculiar algebraic structure: $\hat{\mathbf{F}}$ is not a complete ring, but only a multiplicative group. It contains only a subset of the algebraic numbers: Since there is no addition, the constant of the golden ratio $F^{-1} + F^{-1} 0 1/2$ is not contained. Nevertheless, the factors represented by $\hat{\mathbf{F}}$ cover most known tuning systems.

Relation $\hat{c}$ maps each real number to all of its products with any power of two. This corresponds to factors which determine frequencies which differ only in their *octave register*. In the symbolic world this kind of abstraction is often called *pitch class*; accordingly, each $\hat{c}(X)$ can be called a *frequency class* or *factor class*.

The function c does the same for f-tuples. Table 2 gives two definitions: The second term of the equation tells us to

$$\hat{\mathbf{F}} = \{f : \mathbb{N} \rightarrow \mathbb{Q} \mid f \Vdash \{0\} \text{ finite}\}$$

$$F : \text{SEQ } \mathbb{Q} \leftrightarrow \hat{\mathbf{F}}$$

$$F = \lambda e : \text{SEQ } \mathbb{Q} \bullet (\lambda i : \mathbb{N} \bullet 0) \oplus e$$

$$f : \hat{\mathbf{F}} \leftrightarrow \mathbb{R}$$

$$f = \lambda e : \hat{\mathbf{F}} \bullet \prod \{i : \mathbb{N} \bullet \mathcal{P}(i)^{e(i)}\}$$

$$(*) : \hat{\mathbf{F}} \times \hat{\mathbf{F}} \leftrightarrow \hat{\mathbf{F}}$$

$$(*) = \lambda e, f : \hat{\mathbf{F}} \bullet \lambda i : \mathbb{N} \bullet e(i) + f(i)$$

$$(/) : \hat{\mathbf{F}} \times \hat{\mathbf{F}} \leftrightarrow \hat{\mathbf{F}}$$

$$(/) = \lambda e, f : \hat{\mathbf{F}} \bullet \lambda i : \mathbb{N} \bullet e(i) - f(i)$$

$$(\uparrow) : \hat{\mathbf{F}} \times \mathbb{Q} \leftrightarrow \hat{\mathbf{F}}$$

$$(\uparrow) = \lambda e, f : \hat{\mathbf{F}} \bullet \lambda i : \mathbb{N} \bullet e(i) * f$$

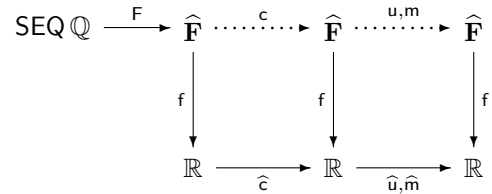
$$\hat{c} : \mathbb{R} \leftrightarrow \mathbb{R}$$

$$\hat{c} = \{x, y : \mathbb{R} \mid \exists z : \mathbb{Z} \bullet x \cdot 2^z = y\}$$

$$\hat{u}, \hat{m} : \mathbb{R} \leftrightarrow \mathbb{R}$$

$$\hat{u} = \lambda x : \mathbb{R} \mid 1 \leq x < 2 \bullet x$$

$$\hat{m} = \lambda x : \mathbb{R} \mid 1/\sqrt{2} < x \leq \sqrt{2} \bullet x$$



$$c, u, m : \hat{\mathbf{F}} \leftrightarrow \hat{\mathbf{F}}$$

$$c = f \sim \circ \hat{c} \circ f = ((*) \circ F)(\mathbb{Z}^1)$$

$$u = f \sim \circ \hat{u} \circ f$$

$$m = f \sim \circ \hat{m} \circ f$$

Table 2. Construction of F-operator.

go from the f-tuple to the factor by applying f , then apply $\hat{c}$, and finally go back again, as depicted in the diagram.

The third term operates directly on f-tuple arguments: Build all singleton sequences of integer numbers, convert these to f-tuples, and multiply them with the arguments.

Both $\hat{c}$ and c are frequently found in harmonic theories, when intervals are discussed under abstraction from the octave registers of the concrete sounds.

The functions $\hat{u}$ and $\hat{m}$ simply filter their arguments and let pass only those factors which falls in the semi-closed range given in the formulas— u and m do the same for f-tuples. Applied to octave classes, they select one factor.

Let $x : \mathbb{R}$ be a frequency factor, then both $u(c(x))$ and $m(c(x))$ select a *special representative* from the factor’s class: It corresponds to the musical interval which is upward and smaller than a perfect octave, or upward or downward of minimal size, respectively.

$$\begin{aligned}
F\ a\ b\ c &= F(\{(a, b, c)\}) = \{F(a, b, c)\} \\
cF\ a\ b\ c &= c(\{F\ a\ b\ c\}) = (c \circ F)(\{(a, b, c)\}) \\
uF\ a\ b\ c &= u(\{cF\ a\ b\ c\}) = \{(u \circ c \circ F)(a, b, c)\} \\
mF\ a\ b\ c &= m(\{cF\ a\ b\ c\}) = \{(m \circ c \circ F)(a, b, c)\}
\end{aligned}$$

$$\begin{aligned}
Fn\ b\ c &= F\ 0\ b\ c \\
cFn\ b\ c &= c(Fn\ b\ c) \\
uFn\ b\ c &= u(cFn\ b\ c) \\
mFn\ b\ c &= m(cFn\ b\ c)
\end{aligned}$$

$$F\ a\uparrow b\ c\uparrow d = F\ \underbrace{0\ 0\ \dots\ 0}_{\mathcal{P} \sim (a)}\ b\ \underbrace{0\ 0\ \dots\ 0}_{\mathcal{P} \sim (c)-1-\mathcal{P} \sim (a)}\ d$$

$$u(r : \mathbb{R}) = \hat{u}(\hat{c}(\{r\})) = \{(\hat{u} \circ \hat{c})(r)\}$$

$$m(r : \mathbb{R}) = \hat{m}(\hat{c}(\{r\})) = \{(\hat{m} \circ \hat{c})(r)\}$$

Table 3. Abbreviated writing of F-operator functions.

Thus $u(c(F\ 0\ 1))$ corresponds to the pure fifth upwards and $m(c(F\ 0\ 1))$ the pure fourth downwards.

2.3 Shortcuts; Abbreviated Writing and Pronunciation

Additional rules concerning the written and spoken representation of all these functions aim at a compact and specific appearance, see Table 3:

1. Each nesting of function names from F-operator can be written and spelled as collapsed into one single function name.
2. When embedded in prose text, the argument list of all functions can be written and spoken without framing parentheses, and with simple whitespace as separators. (The latter has been inspired by [9].)
3. The F operator binds weaker than the division, so that the simple sequence $uFn\ 1/2\ 1/13$ means two rational exponents. More complex expressions must use parentheses, as in $uFn(1/13*r)$ or $(uFn\ 1/13)*r$.
4. In that abbreviated notation, the very first tuple component can be the predefined constant n . This is sensible to indicate that the octave value is “don’t care”, because in any case the tuple shall be argument to the c function. This n is also collapsed into the compound function word, which gives the frequently used constructor cFn .
5. When the context requires a number, the function f is applied implicitly to singleton values from $\hat{\mathbb{F}}$.
6. Because u and m select from classes, the class building operation c is applied implicitly.
7. Higher exponents can be named explicitly instead of by position. For instance $F\ 11\uparrow b\ 13\uparrow d$ allows to speak about powers of the mentioned primes without the need to count a non-trivial prefix of zeros.

The last two lines in Table 3 make u and m directly applicable to factors instead of f -tuples, by overloading. Therefore $u\ 5$ and $m\ 3$ can be used to construct the pure third and pure forth, respectively.

2.4 Evaluating to Rationals

Restricting the arguments of F to integer values yields the restricted functions

$$\bar{F}, \bar{uF}, \bar{mF} : \text{SEQ } \mathbb{Z} \rightarrow \mathbb{Q}$$

Sequences of integer exponents can thus be evaluated to a rational number. Whenever two such rationals are treated as pitches, by applying them to the same reference pitch and frequency, their arithmetic quotient stands for the interval between those pitches.

Whenever two such rationals stand for two intervals, one often wants to compare their sizes. A usual method is to take the rational values written as fractions, convert them to the least common denominator, and compare the numerators. Then the ratio of the numerators corresponds to the interval between the intended pitches. For instance, it turns out that four natural fifths $= uFn\ 4 = 81/64$ are larger than one natural third $= uFn\ 0\ 1 = 5/4 = 80/64$, and that the interval between them, the “syntonic comma”, is $81/80$. Written completely in F-operator, it holds that $u(Fn\ 4 / Fn\ 0\ 1) = uFn\ 4 - 1$.

Another method of comparing is to write the rational results as *decimal fractions*. These are nothing else than compound numbers adding an optional integer part to the left of the decimal dot, an optional fractional part with a denominator 10^k , plus an optional periodically infinite fractional part with a denominator $10^m * (10^n - 1)$, with $k, n : \mathbb{N}_1$ and $m : \mathbb{N}$. This way of writing also makes interval sizes obvious. But because of the somehow arbitrary base “10”, most figures do no longer indicate the underlying construction. The syntonic comma becomes 1.0125, and its size can be compared to, for instance, the Pythagorean comma $uFn\ 12 = F - 19\ 12 = 531441/524288 = 1.013643..$

Comparing both intervals in the F-operator notation is done by division: $F - 19\ 12 / F - 4\ 4\ -1 = F - 15\ 8\ 1 = 32805/32768 = 1.00113..$

2.5 Evaluating to Reals

If and only if there is a non-integer rational among the exponents, the resulting frequency factor is not rational any more. As a consequence, there is no simple number format to notate it, and also in common computer systems most representations are only approximations. While this restriction is almost never relevant in signal-processing practice, it is important to keep it in mind for a correct understanding of the nature of these intervals.

Furthermore, for to get an intuitive feeling for interval sizes, for instance when singing, it is desirable to have in the world of exponents, which determines our perception [1, ?, p. 10] some unifying mechanism which allows easy and intuitive comparison, as the common denominator does in the world of factors.

Common practice is to convert the exponential representations which involve several different bases to one single exponent of the single basis two. Thus a common measure for interval size is $f(F(1/12 * 1/100))$, which is called the *cent interval*. The postfix unit “cent” is append to logarithms of frequency factors. To convert cent values back to factors, an exponential function must be applied. For

	$\log _ \in \mathbb{Q} \text{ cent}$	$\log _ \notin \mathbb{Q} \text{ cent}$
$F(-) \in \mathbb{Q}$	$F 2 = 4 = 2400 \text{ cent}$ (double octave)	$\text{uFn } 0 \ 1 = 5/4 = 386.314 \dots \text{ cent}$ (third in just intonation)
$F(-) \notin \mathbb{Q}$	$F(1/2) = \sqrt{2} = 600 \text{ cent} = 1.41421 \dots$ (half octave)	$\text{uFn}(1/13) = \sqrt[13]{3} = 1.0881822 \dots = 146.3042 \dots \text{ cent}$ (Bohlen–Pierce tempered half-tone) $\text{uF}(1/2) \ 2 = \sqrt{2} * 9/8 = 1.5910 \dots = 803.9100 \dots \text{ cent}$ ("formed" $G^\sharp$ from [3], see sect. 3.6)

Table 4. Examples for combinations of rational and irrational factors and Cent values.

isolated values, this conversion can be left implicit, see Table 4. In practice, often decimal fractions are notated followed by this unit, to approximate real numbers.

2.6 Rational and Irrational Categories of f-tuples

According to the theorem of Lindemann and Weierstraß [11], those cent value of factors $\neq 1$ with other bases than 2 cannot be rational. The conversion to cent values is frequent in practice and indispensable for prima-vista intuition for any reader, for didactic demonstrations, for experiments in psychology and physiology, and many more.

But in harmonic theory it is hardly useful, since it is (nearly always) only an approximation, which does not explain the nature (Hegel: "Wesen") of an interval, but obscures it.

The authors have spent considerable time and effort in correcting the German Wikipedia pages. What frequently has to be done is (a) prepend the text "approximated" to table headers, (b) write the cent values with trailing dots, and/or (c) replace "=" by "≈".

This is tedious, but necessary. Especially laymen and beginners must be protected from the false impression that music intervals are given their characteristics by some "mystical method of quantification". Indeed, it is by comparatively simple *qualitative* derivation.

As a consequence, all f-tuples fall in one of four categories: The factor is rational and the cent value is rational, if and only if there is only one integer exponent, and this is for the base two. (This exponent may be zero.)

The factor is irrational and the cent value is rational, if and only if there is only one non-integer rational exponent and its base is two.

The factor is rational and the cent value is irrational, if and only if only integer exponents are included, and at least one base other than two.

The factor is irrational and the cent value is irrational, if and only if at least one non-integer exponent and one base other than two are included.

Table 4 shows these rules in formulas and typical examples for all cases. Every computer implementation of F-operator must make very clear which kind of numbers is *meant*, independently of the implementing data type of the employed language. In this article we use the format "1.234.." with two dots for (voluntary) abbreviations of rational numbers, but "1.234..." with three dots for (necessary) abbreviations of irrational numbers.

3. EXAMPLE APPLICATIONS

3.1 Deriving the pure fifth from the third partial tone

Nearly all theories on common Western harmonics have the *triad* as their basic material, and derive its structure from that of the spectrum of partial tones, the harmonic series. The frequency factors to be applied to the base frequency are $F \ 1$, $F \ 0 \ 1$, $F \ 0 \ 0 \ 1$, $F \ 0 \ 0 \ 0 \ 1$, etc.

When we define all harmonic effects to happen "modulo octave"—which certainly is a simplification, but a starting point—then $F \ 0 \ 1$ is the first interval from the list above which generates harmonic content. From this the *just fifth* is derived, as the framing interval of the triad. This derivation can be achieved either by setting the frequency in relation to the next lower power of 2, which is the interval of the octave. So the just fifth can be derived and written as $F \ -1 \ 1$. Or it can be done by our class construction and picking the smallest representative greater than one by $\text{uFn } 1$. It holds

$$F \ -1 \ 1 = \text{uFn } 1 = 3/2$$

Stacking arbitrary many just fifths, we never reach the frequency class of the starting point, i.e. its frequency modulo octave, again.

$$a \neq b \iff \text{uF}(0, a) \neq \text{uF}(0, b)$$

No power of 2 (except $2^0 = 1$) is also a power of 3; so we get the well-known "spiral of fifths". This constitutes a scale, unlimited to both sides, called *Pythagorean*, as

$$\{z : \mathbb{Z} \bullet \text{uF}(-, z)\}$$

Typical scales for music purposes are contiguous segments from this scale, typically of five to seven steps, see for instance most of the Japanese scales in [12]. A few others described there contain gaps and thus semitones, see "Hirajoshi" and followers in Table 10.

Table 6 shows the first powers of 3, sorted by increasing exponents. After twelve fifths we reach a frequency factor which has a significantly smaller distance to its lower neighbor (in the scale constructed so far) than its predecessors. It thus has a different sound quality and cannot be used for melodic or harmonic purposes in the same way as all the others. This interval is the *Pythagorean comma* $= \text{uFn } 12 = F \ -7 \ 12 = 531441/524288 = 1.0136433 \approx 23.46 \text{ cent}$.

$$\text{arrangedPureMajorTriads} = \{R \subset \text{cF}(0); T \subset \text{cF}(0, 0, 1); F \subset \text{cF}(0, 1) \mid R \neq \emptyset \wedge T \neq \emptyset \wedge F \neq \emptyset \bullet R \cup T \cup F\}$$

$$\text{arrangedEDOMajorTriads} = \{R \subset \text{cF}(0); T \subset \text{cF}(1/3); F \subset \text{cF}(7/12) \mid R \neq \emptyset \wedge T \neq \emptyset \wedge F \neq \emptyset \bullet R \cup T \cup F\}$$

Table 5. Use of F-operator expressions directly in mathematical formulas.

uF 0	=	fF 0	=	1	=	1	=	1	=	0 cent
uFn 1	=	fF -1 1	=	3/2	=	1.5	=	1.5	≈	701.9550 cent
uFn 2	=	fF -3 2	=	9/8	=	1.125	=	1.125	≈	203.9100 cent
uFn 3	=	fF -4 3	=	27/16	=	1.6875	=	1.6875	≈	905.8650 cent
uFn 4	=	fF -6 4	=	81/64	=	1.265625	=	1.265625	≈	407.8200 cent
uFn 5	=	fF -7 5	=	243/128	=	1.8984375	=	1.898438..	≈	1109.7750 cent
...										
uFn 11	=	fF -17 11	=	177147/131072	=	1.35152435302734375	=	1.351524..	≈	521.5050 cent
uFn 12	=	fF -19 12	=	531441/524288	=	1.0136432647705078125	=	1.013643..	≈	23.4600 cent
uFn 13	=	fF -20 13	=	1594323/1048576	=	1.52046489715576171875	=	1.520465..	≈	725.4150 cent
etc.										

Table 6. Examples of usage: The Pythagorean scale and the Pythagorean comma.

The next twelve steps will thus be simply the first twelve, shifted by this “non-interval”. Therefore most constructions of Pythagorean scales end here.

3.2 Find multiple pure fifths which come near an octave

Two sequences of two different iterated intervals (here: octave and pure fifths) which only meet once, come arbitrarily near elsewhere.

This lemma together with the Pythagorean comma found above, has since long been a motivation to find further near-misses of the spiral of fifths and the octaves. uF 53 motivated the Mercator–Tanaka division of the octave by cF 1/53, see section 6.2 below. The cycle up to cFn 360 had already been found in fourth century China and fitted well into Chinese philosophy: “Including the old tones, there are a total of 360, one for each day.” [13, p. 216].

3.3 Intervals derived from higher components of the harmonic series

Nearly all harmonic theories employ in their calculus further intervals derived from the harmonic series. A systematic way to enumerate candidates is to take descending pairs. In most existing theories, their higher value is subject to a particular limit.

The next interesting element is the fifth one = F 0 0 1. Its next lower octave and its immediate predecessor is F 2, so that uFn 0 1 = F -2 0 1 = 5/4. This is commonly used to derive the interval *just major third*.

The ratio to the next predecessor is F 0 -1 1 = 5/3, is a major sixth. The minimal representative mF 0 -1 1 = 5/6 is less than one, the interval goes *downward*. Swapping the two compared frequencies gives an upward interval, which in most theories is used for further construction. So here: mFn -1 1 = F -1 -1 1 = 5/6 < 1, and uFn 1 -1 = mFn 1 -1 = 6/5 = 1.2 = a *minor third* upward.

The seventh partial tone corresponds to the factor 7 = F 0 0 0 1. It is the basis for “septimal” harmonic theories and music practice, as propagated by Martin Vogel and modelled by *Vogel’s Tonnetz*. [14] The ratios to its non-octave predecessors in the harmonic series are the factors

uFn -1 0 1 = 7/6 = 1.1 $\bar{6}$, called “septimal minor third”, and uFn 0 -1 1 = Fn 0 -1 1 = 7/5 = 1.4, sometimes called the “natural tritone”, which is a very unsystematic name, because Fn 0 -1/3 1/3 is not known as a tone.

The ninth partial tone is not a prime, but the product of two stacked fifths. But it is interesting in relation to the root frequency and the just third: uFn 2 is the *greater tone* and uFn -2 1 is the *lesser tone* of just intonation, see below.

3.4 Compare the powers of the higher components of the harmonic series

The nearby meeting points of powers of fifth and octave motivate to look for similar effects involving the higher intervals. This means to search for some mF ... a ... b ≈ 1, with a, b ≠ 0. Indeed we find mFn 4 -1 = uFn 4 -1 = F -4 4 -1 = 81/80 = 1.0125 = the *syntonic comma*: one just third is slightly smaller than four just fifths.

Furthermore, mFn 0 3 = F -7 0 3 = 125/128 = 0.9765625 = three just major thirds are slightly smaller than an octave, the difference is called *lesser diesis*.

Then there is uFn -2 0 -1 = F -5 -2 0 -1 = 64/63 = 1.015873 = the *septimal comma*: One single just seventh downward (which can be considered one septimal whole tone upward) is slightly larger than two just fifths upward.

And finally uFn 1 0 0 1 = fF -5 1 0 0 1 = 33/32 = 1.03125 = 53.2728... cent = the *undecimal comma*: The 11th partial tone is approximately a quarter tone higher than one just fourth up, which is one pure fifth down.

More complicated are the meetings of powers of more than one basis. The most prominent and relevant is that of the (just) minor third Fn 1 -1 and the octave. It holds that four minor thirds are slightly larger than an octave by mFn 4 -4 = F 3 4 -4 = 648/635 = 1.0368 = the so-called *greater diesis*.

3.5 Common Western Tunings

A major scale relative to any root pitch and its frequency, for example

$$\langle C4, D4, E4, F4, G4, A4, B4(, C5) \rangle$$

is constructed in the 12 EDO tuning, when setting C4 ≡ 1, by the sequence of intervals

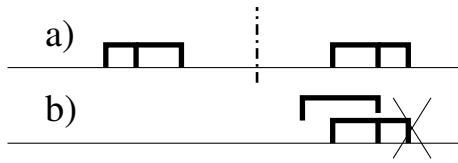


Figure 3. Two different procedures to invert a proportion.

$$\langle F 0, F 2/12, F 4/12, F 5/12, F 7/12, F 9/12, F 11/12 \rangle$$

The Pythagorean construction for the same scale is

$$\langle uF 0, uFn 2, uFn 4, uFn -1, uFn 1, uFn 3, uFn 5 \rangle$$

The *just intonation* sees this scale as a compression of three stacked and overlapping major triads:

$$F - A - C - E - G - B - D$$

It can be derived from the Pythagorean by replacing the thirds of these triads, the pitch classes A, E, and B, by the third as contained in the harmonic series. For these three chords we get

$$\langle uFn -1, uFn -1 1, uFn 0, uFn 0 1, uFn 1, uFn 1 1, uFn 2 \rangle$$

Arranged like the scale at the beginning of this section, we get

$$\langle uF 0, uFn 2, uFn 1 1, uFn -1, uFn 1, uFn -1 1, uFn 1 1 \rangle$$

Written as numbers this is

$$\langle 1, 9/8, 5/4, 4/3, 3/2, 5/3, 15/16 \rangle$$

which is the same as

$$\langle 1, 1.125, 1.25, 1.\bar{3}, 1.5, 1.\bar{6}, 1.9375 \rangle$$

The *derivation of the minor scale* can be done in different ways. When three points on some axis scale are given with two distinct inner distances, the inversion of the proportion can be done (a) by mirroring the whole structure by an arbitrary point of that axis, or (b) by adding the outer distance to the center point of the structure in one direction and discarding the outer point which lies in the other direction. Both procedures switch the order of the inner distances, see Figure 3.

The method (a) is used by the “dualistic” theories of harmonics (early Riemann, Oettingen, Karg-Ehlert), which use an “undertone spectrum” to derive the minor chord. The method (b) is used by the “monistic” theories of harmonics, where the major chords and modes are given, and from these the minors are derived, by re-arranging the patterns on the keyboard.

A monistic way to construct three minor chords and thus a minor scale, is to prepend to the triads above one factor which is a fifth = $uFn 1$ lower than the third of the lowest triad, resulting in a new pitch D^* . So we get the pitch classes of an a-minor scale as

$$D^* - F - A - C - E - G - B$$

which is

$$\langle Fn -2 1, Fn -1, Fn -1 1, F 0, Fn 0 1, Fn 1, Fn 1 1 \rangle$$

This shows clearly that the common textbook sentence “a-minor and C-major use the same set of pitch classes” indeed depends on the applied tuning system: Here the two pitches called D and D^* differ by the syntonic comma.

$$\hat{c}_- : \mathbb{N} \rightarrow (\mathbb{R} \leftrightarrow \mathbb{R})$$

$$\hat{c}_p = \{x, y : \mathbb{R} \mid \exists z : \mathbb{Z} \bullet x \cdot p^z = y\}$$

$$\hat{u}_-, \hat{m}_- : \mathbb{N} \rightarrow (\mathbb{R} \leftrightarrow \mathbb{R})$$

$$\hat{u}_p = \lambda x : \mathbb{R} \mid 1 \leq x < p \bullet x$$

$$\hat{m}_p = \lambda x : \mathbb{R} \mid 1/\sqrt{p} < x \leq \sqrt{p} \bullet x$$

$$c_p = f^\sim \circ \hat{c}_p \circ f$$

$$u_p = f^\sim \circ \hat{u}_p \circ f$$

$$m_p = f^\sim \circ \hat{m}_p \circ f$$

Table 7. A variant of F-operator for systems without octave identity.

The *quarter-comma meantone* tuning can most naturally be derived as follows: When tuning for instance a harpsichord by stacking pure fifths from C2, one reaches with E4 nearly the fifth partial tone of C2. So if we want this partial tone to sound on this key, we restart the whole process: Instead of pure fifths we use the fourth part of its frequency factor = $F 0 0 1/4 = uF 0 0 1/4$. sometimes called “meantone fifth”.

The intonation process follows the spiral of fifths as usual. As with $uF 0 1$, it never comes back to a frequency class it has passed. The quarter-comma meantone tuning thus defines a scale of infinite many frequencies—the historical keyboard tunings realize a segment of twelve of these. The most frequently found keyboard design goes from $E\flat$, $B\flat$, F, ... up to $C\sharp$, $G\sharp$. With $C \equiv 1$, these frequency classes are

$$\begin{aligned} &\langle uFn 0 -3/4, uFn 0 -1/2, uFn 0 -1/4, cF 0, \\ &\quad uFn 0 1/4, cF 0 1/2, uFn 0 3/4, uFn 0 1, \\ &\quad uFn 0 5/4, uFn 0 3/2, uFn 0 7/4, uFn 0 2 \rangle \end{aligned}$$

The *wolf fifth* goes from the last of these to the first: $u(F 0 0 -3/4 / F 0 0 2) = uF 0 0 -11/4 = 1.1993... \approx 738$ cent. It is obviously different from both $uFn 1 = 1.5 \approx 702$ cent and $uFn 0 1/4 = 1.4953... \approx 697$ cent.

The construction of the *third-comma meantone* tuning is one step more indirect: When stacking three pure fifths we come very near to the frequency $F 1 -1 1$, which is a pure fifth lower than the pure third and its class can be called a *minor third* below C (or a major sixth above). So we get $F 1/3 -1/3 1/3$ as the *third-comma meantone fifth*.

Werckmeister III forces the spiral of fifths into a circle by subtracting a quarter of the Pythagorean comma = $F 19/4 3$ from the fifths over C, G, D, and E. The result (with $C \equiv 1$):

$$\begin{aligned} &\langle cF 0, cF -19/4 -2, cF -19/2 -4, cF -57/4 -6, \\ &\quad cF -57/4 -5, cF -19 -7 = cFn -7, cFn -6, cFn -5, \\ &\quad cFn -4, cFn -3, cFn -2, cFn -1 \rangle \end{aligned}$$

3.6 Rare Tunings

Renold et.al. [3, p.95] define a tuning from an anthroposophic background. They consider the equal division of the octave as a consonance. Their tuning is Pythagorean (with pure fifths) for the black and white piano keys separately, both groups connected by $\sqrt{2}$ between C and $F\sharp$. Intervals between black and white keys are called “geformt” (“formed”). The chromatic scale on C thus reads

$$\langle uF 0, uF 1/2 1, uFn 2, uF 1/2 3, uFn 4, uFn -1, uF 1/2, uFn 1, uF 1/2 2, uFn 3, uF 1/2 4, uFn 5 \rangle$$

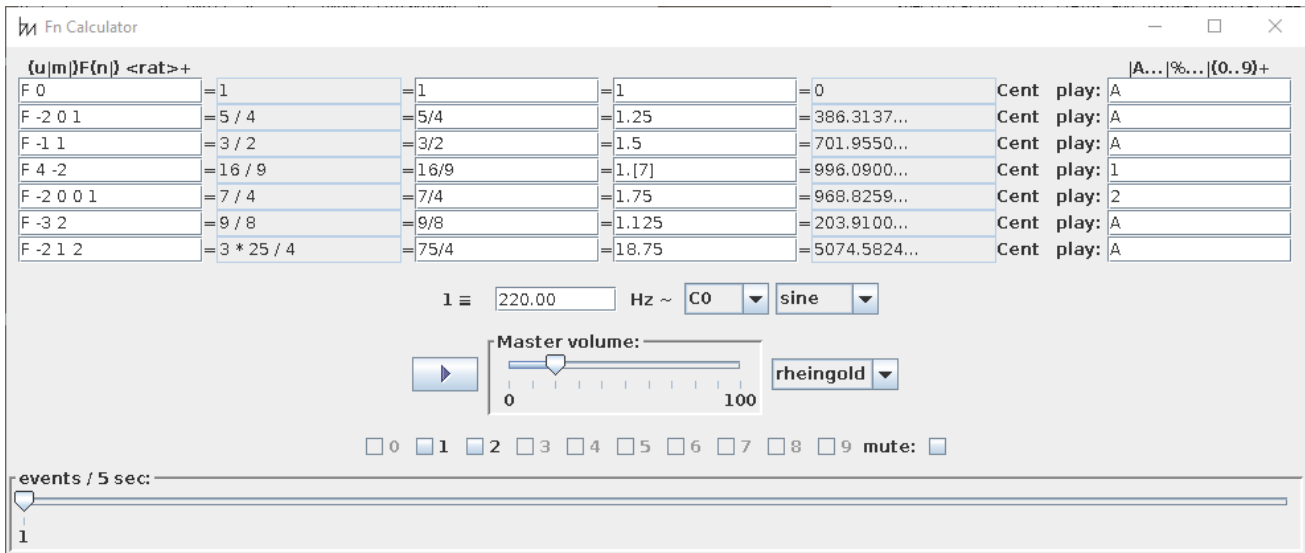


Figure 4. Screenshot of the application to explore F-operator expressions.

3.7 Symbolic calculi not completely compatible with F-operator

The family of *Bohlen–Pierce scales*, and any other system not employing the identity modulo octave, is not fully compatible with all components of F-operator.

A *just* Bohlen–Pierce scale can be described by

$$\langle F0, F_{n-12-1}, F_{n-20-1}, F_{n-0-11}, F_{n-11}, F_{n-2-1}, F_{n-11-1}, F_{n-101}, F_{n-22}, F_{n1} \rangle$$

The intervals between the steps of this scale are easily calculated as the series of differences:

$$\langle F_{n-12-1}, F_{n-3-2}, F_{n-2-12}, F_{n-12-1}, F_{n-3-2}, F_{n-12-1}, F_{n-2-12}, F_{n-12-1}, F_{n-3-2} \rangle$$

How conveniently the F-operator notation can be used to *construct* these scales is still to be explored.

The operators cF, uF, and mF cannot be applied, because they work modulo octave. Table 7 shows more general definitions of the fundamental operators—further definitions can be adopted accordingly.

Any *tempered* Bohlen–Pierce scale can be described like

$$\langle F0, F_{n/13}, F_{n/13}, F_{n/13}, F_{n/13}, F_{n/13}, F_{n/13}, F_{n/13}, F_{n/13}, F_{n/13}, F_{n/13} \rangle$$

4. IMPLEMENTATION

An irrational number which is an instance of $\hat{F}$ can thus be precisely represented in a computer. This is employed by the demo software, see figure 4. The binary is found at <http://bandm.eu/downloads/FnCalc.jar>; the sources shall be published in near future.

Each line stands for one particular factor: The first column allows the input of all F-operator expressions types. These will be normalized to the explicit form “F...”. The second column shows a “fraction of products of powers”, while the third column shows the fractional value, if it is

rational. This column can also be used for input. The next column shows the decimal fraction representation, either exact for rational values or only a prefix, as usual for irrational values. Rational values can be entered. Then follows the output of the cent value. Whenever a value is entered, all other columns are updated accordingly, or the entry field is colored red to indicate a syntax error.

The right half of the GUI makes the factors audible, relative to a user-defined reference frequency: They can be switched on individually and can be arranged to chords, which can be played by a primitive sequencer.

One selection menu is for the kind of sound used for all chords, another initializes one of several demo set-ups.

5. RELATED WORK

Independent and earlier are the proposals by Joe Monzo [9, 10]. Main deficiencies: they have never been scientifically published (“my book” mentioned in [10] is unavailable), the necessary mathematics are not fully developed, and the functions of F-operator beyond the basic “F” are missing. Nevertheless the fundamental idea is very similar, and we gratefully borrow the trick to use simple whitespace as separators for the F-operator abbreviated format.

6. REPHRASING RELEVANT WORKS FROM LITERATURE

6.1 Examples from Wikipedia

All entries in the German Wikipedia related to intervals, tuning, harmonic theory, etc., suffer from the fact that many contributors want to present their personal opinion as mandatory. But indeed in all these fields there are multiple methods, more or less oriented towards practice or towards theory, and a lexicon is hardly suited to present them all, with equal rights and cleanly separated.

As a consequence, we find an overwhelming amount of redundancy, in three concerns: (a) different users “occupy”

[page 8: “Zunächst berechnen wir das Intervall ...]

First we calculate the interval between $\underline{b}$ and $c\flat$: $\underline{b}$ is the pure major third of g , so $\underline{b}/g = \text{uFn } 01$.

You get $c\flat$ by eight steps of a fourth:

$$g - c - f - b\flat - e\flat - a\flat - d\flat - g\flat - c\flat$$

so $c\flat/g = \text{uFn } -8$ and $\underline{b}/c\flat = \text{uFn } 81 = F-1581 = 32805/32768 = S \approx 887/856$.

This interval, called “schisma”, is according to W. Freyer already beyond perceptive distinction. [...] So we get the following equivalences (with uFn as distances from g):

$$\underline{b} = \text{uFn } 01 \equiv c\flat = \text{uFn } -8 \text{ and } \underline{\underline{a}}\sharp = \text{uFn } 12 \equiv \underline{b}\flat = \text{uFn } -71 \text{ and } f\sharp = \text{uFn } 5 \equiv \overline{g}\flat \text{ [sic!]} = \text{uFn } -3-1 \text{ (etc...)}$$

[page 9: “Ferner wollen wir das Intervall...”]

Furthermore, we want to compare a tone from the diagonal row of steps of thirds with a nearest tone from the neighboring diagonal. Let us compare $\overline{\overline{f}}\flat$ and $\underline{\underline{e}}\sharp$.

From $\overline{\overline{f}}\flat$ to $\underline{\underline{a}}\sharp$ downwards by minor thirds is

$$\overline{\overline{f}}\flat - \overline{d}\flat - \overline{b}\flat - g - e - c\sharp - \underline{\underline{a}}\sharp$$

We get the factor $\overline{\overline{f}}\flat/\underline{\underline{a}}\sharp = \text{uFn } -66$. But $\underline{\underline{a}}\sharp/\underline{e}\sharp = \text{uFn } -1 = 3/4$.

Thus $\underline{\underline{e}}\sharp/\overline{\overline{f}}\flat = \text{uFn } -66 * \text{uFn } -1 = \text{uFn } -56 = F-656 = 15625/15552 = 1.004694.. = K \approx 214/213$.

This interval I want to call “kleisma”, which in higher ranges can only be heard with dedicated concentration. One can therefore set without losing the tuning (uFn again measured from g):

$$\overline{\overline{g}}\flat = \text{uFn } 5-3 \equiv \underline{\underline{f}}\sharp = \text{uFn } 3 \text{ and } \overline{b}\flat = \text{uFn } 5-2 \equiv \underline{\underline{a}}\sharp = \text{uFn } 04 \text{ etc.}$$

[page 10: “Durch Anwendung der schismatischen und kleismatischen Verwechselung ...”] By applying the schismatic and kleismatic changes, we can reduce the number of distinct tones to 53, see Figure 5.

[page 15: “Von einem Tone ausgehend ...”]

Starting with a particular tone, go three fifths to the right and a minor third down, i.e. add $\mathfrak{R} = \text{uFn } 4-1$. You reach a tone which is a syntonic comma sharper than the start point. Repeat this until a cycle is closed. Whenever you cross the border into the neighboring parallelogram which is labeled $+K$, this indicates that all pitches here are one K sharper than in the main parallelogram. Into this we jump back, so we execute a smaller step than a normal $\mathfrak{R}$, so we have to subtract K from the reached frequency. Correspondingly for $-S$, etc.

For instance, start from $c = \text{uFn } 0$, follow three fifths to the right to $a = \text{uFn } 3$ and then a minor third up to $\bar{c} = \text{uFn } 4-1$.

The next such step goes in fifths over the border to $\overline{\overline{b}}\flat$ and over the second border to $\underline{\underline{c}}\sharp$. The tone $\underline{\underline{c}}\sharp$ is found again in the diagram, in its top left corner, and we jump there. This calculates to $u(\text{uFn } 4-1 * \mathfrak{R} * K/S) = u(\text{uFn } 4-1 * \text{uFn } 4-1 * \text{uFn } -56 / \text{uFn } 81) = \text{uFn } -53$

From there we can continue to $\underline{\underline{c}}\sharp = \text{uFn } -12$ without further border crossings.

Every single step in a complete cycle spans either an altered or a normal syntonic comma, whether or not a border is crossed:

- twelve steps cross only the right border and thus include one schismatic change. Its size can be labelled $\mathfrak{R}/S = \text{uFn } 4-1 / \text{uFn } 81 = \text{uFn } -4-2$.
- five steps cross only the lower border and thus contain a kleismatic change. Their size is $\mathfrak{R} * K = \text{uFn } 4-1 * \text{uFn } -56 = \text{uFn } -15$.
- seven steps contain both changes, of size $\mathfrak{R} * K/S = \text{uFn } 4-1 * \text{uFn } -56 / \text{uFn } 81 = \text{uFn } 74$.
- 29 steps of unaltered syntonic commata of size $\mathfrak{R}$.

By executing this whole cycle of 53 steps of (normal and altered) syntonic commata, we finally reach $\mathfrak{R}^{29} * (\mathfrak{R}^{12}/S^{12}) * (\mathfrak{R}^5 * K^5) * (\mathfrak{R}^7 * K^7/S^7) = \mathfrak{R}^{53} * K^{12}/S^{19} = F-212212-53 * F-72-6072 / F-28515219 = F1 = 2$.

It follows that $k = 2/\mathfrak{R}^{53} = K^{12}/S^{19} = F-72-6072 / F-28515219 = F213-21253 = 1.03536.. \approx 29/28$ is the factor by which a sequence of 53 unaltered syntonic commata is too small to close the octave.

So we can close the cycle by a modified step width of $(\text{uFn } 4-1) * \sqrt[53]{k}$.

This division can much more directly be realized by $F1/53$, as practiced since long (Didymus(?), Kircher, N. Merkator). But with our derivation, the approximation of the just third $F(17/53) \approx \text{uFn } 01$ is not merely accidental but intended.

Table 9. Rephrasing of the derivation of 53 EDO starting from just intonation according to Tanaka [15]. See Figure 5.

is supported by guitar fret diagrams and keyboard diagrams, given for every scale. Its intentions are doubtlessly laudable. But it is questionable to quote scale constructions from overall of the world and transfer them with only little comment to the 12 EDO keyboard.

Late in the text, Hewitt confesses “[In case of] Indonesian Gamelan, [...] Thai, [...] and African [...] music, tuning is more or less an indispensable feature of the sound. Consequently to try and bring these scales over and adapt them to the Western scale is often to lose a great part of their original charm.” [section 9, p. 140]

But this applies also to all other scales, for which the underlying tuning is merely given by a label, like “Tuning: Twelve-tone equal temperament” or “Tuning: Just”. which may be ambiguous with particular scales, or even less un-specific: “Tuning: By ear” or “Tuning: Varies.”

Already for the first third of the book, the more or less conventional Western scales, the tuning indication is often well-nigh false: It must be doubted that a solo violinist playing in “Hungarian minor I” uses 12 EDO intonation—the same with Gregorian singers using Dorian mode.

Only sections 9 to 11 additionally contain cent values for the reader to tune their synthesizer. For gamelan music, he quotes *measurements* carried out by Ellis, McPhee, Kunst, and others, see also p. 159. But in the context of African “Silaba” and “Tomora Masengo” [p. 165f.] just intervals (empirically measured) are in the same scale with cent values 500 and 700 (obviously from 12 EDO).

We suppose that this violates the fundamental difference between cent values coming from measured physical reality, and tuning systems, which are *concepts* and can be found by interviewing musicians and instrument makers.

The merits of this book would be much larger when the scales had been given in their true, intended structure, throughout. To *approximate* this by a 12 EDO keyboard is a subsequent step. That the first step has been skipped is understandable w.r.t. the addressed audience, but nevertheless a regrettable methodological flaw.

For all intended structures, the F-operator system would have been a better choice for presentation, as shown in table 10 for some examples. They start with the acoustic scale, because there the criticized deficiency is already indicated by its name. The laconic label “tuning: just” is for many scale degrees ambiguous (see “Bilawal Thāt”, p. 161, and “Mela Naganandini”, p. 209) and also can be clarified by simple F-operator expressions.

6.4 Free combination of tuning systems in *bach* for Max

The *bach* system is a very elaborate library for compositional operations (in a traditional sense of the word) in the context of the Max framework for processing. Agostini and Ghisi [16] present an extension of the pitch notation in this library, suited to support *just tuning*.

Before this extension, the notation was already overloaded: each English term for a pitch (i.e. a pitch class name followed by the number of an octave register) stands either for the frequency of this pitch, or for the factor of this frequency relative to that of the pitch “C0”. These frequencies

and factors are calculated by applying the 12 EDO tuning to the pitch names.

A new additional format for pitches is now written with an inserted pair of braces, like “C{ }4”. These expressions stand for the frequencies and factors resulting from *Pythagorean* tuning, following the spiral of fifths.

To describe scales in *pure tuning in all keys*, the count of the *commata* to apply is inserted into the braces. They are identified by their position, separated by semicola, and trailing zeroes are of course implicit. Using explicit *commata* is a very established method, see for instance Helmholtz [2, p. 453] for the syntonic comma. It has the disadvantage that their direction is not evident without re-calculating their construction—does the “11-comma” go up or down in the frequency world? Indeed, using *commata* to define pitches is a *detour*: First we must notate what we do *not* mean, and then apply the necessary correction.

The big advantage realized by this extension to *bach* is its *compositionality*: Both kinds of frequencies and factors can freely be mixed in all arithmetic operations. Agostini and Ghisi point out that a 12-tone-tempered frequency can be added to and subtracted from any just interval, etc.

Indeed this can have practical relevance: Hitting an E2 fortissimo on a grand piano, and then filtering the resulting sound for its fifth partial tone needs a filter frequency which is such a mixture. In *bach* that task can now be written as $E2 + E\{-1\}2$: Start with the 12 EDO tone E2, add four just fifths (which happens to be a Pythagorean major third plus two octaves, thus written $E\{2\}$), and then go down one syntonic comma again, because you don’t want the Pythagorean but the just third. Together with the two octaves, this third reaches the fifth partial tone. In F-operator we can rephrase as $F(2 + 4/12) * F\ 0\ 0\ 1 = F\ 7/3\ 0\ 1$.

The presentation of paper [16] on the Tenor 2025 conference has motivated the development of F-operator.

7. ACKNOWLEDGEMENTS

Thanks to all anonymous reviewers for valuable suggestions to improve this article.

8. REFERENCES

- [1] C. Stumpf, *Die Anfänge der Musik*. Leipzig: J. A. Barth, 1911, [https://www.digitale-sammlungen.de/en/view/bsb00115263\[20260114\]](https://www.digitale-sammlungen.de/en/view/bsb00115263[20260114]).
- [2] H. von Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Friedrich Vieweg und Sohn, 1896. [Online]. Available: <https://books.google.de/books?id=-osAAAAYAAJ&pg=PR3>
- [3] M. Renold, *Von Intervallen, Tonleitern, Tönen und dem Kammerton $c = 128\text{ Hz}$* . Dornach: Philosophisch-Anthroposophischer Verlag am Goetheanum, 1985.
- [4] M. Vogel, *On the Relations of Tone*. Bonn: Verlag für systematische Musikwissenschaft, 1993.
- [5] M. Hewitt, *The Tonal Phoenix*. Bonn: Orpheus Verlag, 2000.

Section 3: Synthetic Modes and Scales							
acoustic	uF 0	uFn 2	uFn 0 1	uFn 11 ↑ 1	uFn 1	uFn 13 ↑ 1	uFn 0 0 1
Hungarian minor I	uF 0	uFn 2	uFn 1 -1	uF 2 1	uFn 1	uFn 0 -1	uFn 2 -1
						(or	uFn 0 0 1 ?)
Hungarian minor II	uF 0	uFn 2	uFn 1 -1	uF 2 1	uFn 1	uFn 0 -1	uFn 1 1
Double harmonic	uF 0	uFn -1 -1	uFn 0 1	uF -1	uFn 1	uFn 0 -1	uFn 1 1
Section 7: Pentatonic Scales of Japan							
Ryo	uF 0	uFn 2	uFn 4	uF 1	uFn 3		
Ritsu	uF 0	uFn 2	uFn -1	uF 1	uFn 3		
Min'yō	uF 0	uFn -3	uFn -1	uF 1	uFn -2		
Hirajoshi	uF 0	uFn 2	uFn -3	uF 1	uFn -4		
Iwato	uF 0	uFn -5	uFn -1	uF -6	uFn -2		
Kokin Joshi	uF 0	uFn 2	uFn -1	uF 1	uFn -4		
Akebono	uF 0	uFn 2	uFn -3	uF 1	uFn 3		
Yosenpo	uF 0	uFn 2	uFn -1	uF 1	↗uFn 4	↘uFn -2	
Insenzo	uF 0	uFn -5	uFn -1	uFn 1	↗uFn -2	↘uFn -4	
Section 9: Indonesian Scales of the Gamelan							
Slendro	uF 0	uF 1/5	uF 2/5	uF 3/5	uF 4/5		
Section 10: Equatonal Scales of Thailand							
Thai seven-toned	uF 0	uF 1/7	uF 2/7	uF 3/7	uF 4/7	uF 5/7	uF 6/7
Thai mode 1	uF 0	uF 1/7	uF 2/7		uF 4/7	uF 5/7	
Thai mode 3	uF 0	uF 1/7		uF 3/7	uF 4/7		uF 6/7
Section 12: Scales of Northern India							
Khamaj Thāt	uFn 0	uFn 2	uFn 0 1	uFn -1	uFn 1	uFn 3	uFn -2
	(or	uFn -2 1				uFn -1 2	uFn 2 -1 ?)
Section 13: Scales of Carnatic Music							
Mela Naganandini	uFn 0	uFn 2	uFn 0 1	uFn -1	uFn 1	uFn 2 2	uFn 1 1
					(or	uFn 6 1 ?)	

Table 10. Some proposals to enhance “Musical Scales of the World” [12] by substantial tuning information.

- [6] R. Ploeger, *Studien zur systematischen Musiktheorie*. Eutin: Petersen Mikkelsen, 2002.
- [7] J. Haluška, *The Mathematical Theory of Tone Systems*. New York, Basel: Marcel Dekker, 2004.
- [8] M. Lepper, B. T. y Widemann, and M. Oehler, “fun-code – versatile syntax and semantics for functional harmonic analysis labels,” *Music & Science*, 2022, <https://doi.org/10.1177/20592043221085659>.
- [9] J. L. Monzo, “monzo / exponent vector / prime-exponent vector,” *tonalsoft*, 2005. [Online]. Available: <http://www.tonalsoft.com/enc/m/monzo.aspx>
- [10] —, “Justmusic prime-factor notation,” *tonalsoft*, 1997. [Online]. Available: <http://tonalsoft.com/monzo/article/article.htm>
- [11] A. Baker, *Transcendental Number Theory*, London, 1975.
- [12] M. Hewitt, *Musical Scales of the World*. The Note Tree, 2013. [Online]. Available: <https://pdfcoffee.com/musical-scales-of-the-world-pdf-free.html>
- [13] E. G. McClain and M. S. Hung, “Chinese cyclic tunings in late antiquity,” *Ethnomusicology*, vol. 23, no. 2, pp. 205–224, 1979.
- [14] M. Vogel, *Die Lehre von den Tonbeziehungen*. Bonn–Bad Godesberg: Verlag für systematische Musikwissenschaft, 1975.
- [15] S. Tanaka, “Studien im Gebiete der reinen Stimmung,” *Vierteljahresschrift für Musikwissenschaft*, vol. 6, no. 1, pp. 1–90, 1890. [Online]. Available: https://books.google.de/books?id=c485AAAAIAAJ&pg=PA1&redir_esc=y#v=onepage&q&f=false
- [16] A. Agostini and D. Ghisi, “Just intonation in bach,” in *Proceedings of the International Conference on Technologies for Music Notation and Representation – TENOR’25*, Beijing, 2025. [Online]. Available: <https://www.tenor-conference.org/proceedings.html>
- [17] J. M. Spivey, *The Z Notation: A reference manual*, ser. International Series in Computer Science. Prentice Hall, 1988, [https://spivey.oriel.ox.ac.uk/corner/Z_Reference_Manual\[20240729\]](https://spivey.oriel.ox.ac.uk/corner/Z_Reference_Manual[20240729]).

A. MATHEMATICAL NOTATION

The employed mathematical notation is fairly standard, inspired by the Z notation [17]. The following table lists some details:

$\mathbb{N}$	All natural numbers, incl. Zero.
$\mathbb{Z}, \mathbb{Q}$	All integer and rational numbers, resp.
$A \times B$	The product type of two sets A and B .
$A \leftrightarrow B$	The type of the relations between A and B .
$A \rightarrow B, A \rightharpoonup B$	The total and partial functions from A to B .
$r \sim$	The inverse of a relation
$f \downarrow s \downarrow$	The image of the set s under the relation f
$r \circ s$	The composition of two relations.
$\text{dom } A, \text{ran } A$	Domain and range of a relation.
$r \oplus s$	Overriding of function or relation r by s .
$R \triangleright S$	$= R \setminus (\text{dom } R \times S)$, i.e. negative range restriction of a relation.
$\text{SEQ } A$	Lists as maps $\{0..n\} \rightarrow A$.
$\langle \dots \rangle$	Instance of a list type.

Functions are considered as special relations; relations as sets of pairs. So with functions, expressions like “ $f \cup g$ ” are defined.

A PROSODIC LAYER FOR SUPRASEGMENTAL FEATURES SPANNING RHYTHM TREES, MOTIVATED BY CARNATIC RHYTHMICAL STRUCTURES

Daniel Miller

Independent Researcher

danielmiller.js@gmail.com

ABSTRACT

Rhythm trees model durational structure as recursive equal subdivision, but musical rhythm often involves suprasegmental patterns that cross-cut the subdivision hierarchy—long-distance dependencies that context-free grammars cannot express. Carnatic rhythmic techniques provide clear test cases: *rhythmic sangati* preserves motives across changes in subdivision density and accent placement, while *jathi bhedom* treats irregular accent spans as compositional units.

Drawing on work in prosodic phonology which models intonation as a tree-based hierarchy distinct from segmental structure, we introduce an *accent tree* that overlays a base rhythm tree. The accent tree acts on a depth projection—a sequence of positions at a chosen metrical level—and partitions those positions into spans that may cross subdivision boundaries. Each span either inherits the underlying rhythm or hosts an attached rhythm tree time-scaled to fill it. Accent trees compile to labeled duration sequences that map back to rhythm trees, so the formalism remains compatible with existing rhythm-tree grammars while capturing long-distance dependencies.

1. INTRODUCTION

Rhythm trees provide a compact, context-free account of durational structure as recursive equal-subdivision of time. More precisely, they are *duration* trees: they encode how time is split, not how the resulting units are grouped, stressed, or otherwise made salient. Yeston [1] observed that rhythm, as perceived, involves criteria beyond duration: timbre, dynamics, density, and pattern recurrence help pick out groupings not derivable from subdivision alone. For Yeston, however, these groupings result from interacting periodic strata, not as independent structures in their own right. The Carnatic techniques developed in Section 3 suggest a need for more flexible accentual structures that may form their own hierarchies, unfolding over extended spans of musical time and crossing durational boundaries.¹

¹ Reflecting the term’s prominence in the Carnatic theory literature, in what follows we use the term “accent” as a cover term for any rhythmic grouping induced by non-durational features such as timbre, dynamics, or orchestration, imposed on a durational grid that partitions the same

Just as linguistic prosody organizes words into phrases that may cross-cut syntactic boundaries, we argue that an analogous separation is needed for formal models of rhythm: a *duration–accent interface* that treats these as distinct but related structural layers.

Carnatic techniques like *jathi bhedom* and rhythmic *sangati* encode musical organization through accentual patterns that have no clear representation as rhythm trees. Recent work has extended context-free rhythm grammars to handle syncopation and anticipation through local mechanisms such as timestealing [2], but the Carnatic cases involve accent constraints that span entire cycles and cannot be localized to adjacent timespans. This is the kind of non-locality for which prosodic phonology introduced a separate structural layer in linguistics, and we argue that the same separation is needed for rhythm.

In prosodic phonology and speech technology, tree based models of intonation treat prosodic structure as a separate hierarchical representation related to, but not reducible to, syntactic structure or segmental strings [3, 4]. Prosodic trees in this tradition organize words or syllables into constituents at several levels, and every node in the tree is interpretively active for prediction and classification tasks. The prosodic hierarchy is anchored in syntax and segmental timing, but it has its own topology and supports systematic patterns of alignment or misalignment between levels.

We propose an analogous architecture for musical rhythm: a two layer model in which a rhythm tree provides an equal subdivision grid and a separate accent tree captures the hierarchical grouping of metrical positions into accented spans, independent of their internal durational content. The accent tree operates over a *depth projection* of the rhythm tree whose groupings may cross-cut the subdivision hierarchy from which its constituent positions are drawn. In line with tree based models of prosodic structure in speech, we treat every node in the accent tree as interpretively active: each node defines a potential accent domain, and the accentual status of any surface event is determined by the lowest accent node that dominates it. This provides a simple formal account of the accentual clash between *gati* and *jathi*. *Gati* fixes a durational grid and suggests one accentual pattern at the subdivision level; *jathi* induces a potentially conflicting pattern that groups those same positions into spans of different cardinality. Both are represented in the accent tree, and where they conflict, the *jathi* grouping takes precedence.

temporal span as the subdivision hierarchy but is not derived from it.

Although the formal motivation comes from Carnatic rhythmic theory, the structures it identifies are not confined to that tradition. Figures 2 and 3 reconstruct passages from Elliott Carter’s *Eight Pieces for Four Timpani* using the accent-tree formalism, illustrating how Carnatic-like structural elements—cross-cutting accent groupings, recursive subdivision of accent spans, and metric modulation as a consequence of hierarchical repartitioning—arise independently in twentieth-century concert music.

The remainder of the paper proceeds as follows. Section 2 reviews rhythm trees and their rewrite theory. Section 3 summarizes Carnatic rhythmic organization and distills the formal challenges these techniques pose for rhythm-tree representations. Section 4 defines accent trees over depth projections, introduces the attachment model with recursive and iterative modes of application, and shows how rhythmic sangati arises naturally as a compatibility condition between accent structures sharing equal capacity. We conclude with a discussion of directions for future work, including generative grammars for accent trees and computable measures of rhythmic tension.

2. BACKGROUND: RHYTHM TREES

A rhythm tree represents a single span of time that is recursively subdivided into equal parts. Each internal node specifies that its duration is split into p equal segments, where $2 \leq p \leq P$ for some fixed maximal arity P .

Let $\Sigma = \{n, _ \}$ be the alphabet of terminal symbols, where n marks the onset of a new event and $_$ marks a tie that prolongs the most recent onset. A rhythm tree over Σ is defined recursively as either a leaf $\ell \in \Sigma$ or an internal node of the form

$$t = p(t_1, \dots, t_p), \quad (1)$$

where $2 \leq p \leq P$ and each t_i is itself a rhythm tree. Following Jacquemard et al. [5], we label each internal node with its arity. The *depth* of a node is its distance from the root, with the root at depth 0.

We assume the root of every rhythm tree spans a duration of 1. Each internal node of arity p splits its incoming duration equally among its children: if a node has duration D and arity p , each child receives duration D/p .

Let $(\ell_0, \dots, \ell_{L-1})$ be the leaves of t in left-to-right order, with corresponding durations $(d_0, \dots, d_{L-1})$. The *labeled duration sequence* of t is

$$\text{lds}(t) = ((\ell_0, d_0), \dots, (\ell_{L-1}, d_{L-1})). \quad (2)$$

Two rhythm trees are *equivalent* when they have the same labeled duration sequence.

Jacquemard et al. [6] give a rewrite system for rhythm trees that is sound and complete: equivalent trees are connected by rewrite paths, and every rewrite preserves rhythmic value. Crucially, each rewrite step is local to the subdivision hierarchy: rules can normalize or repartition existing constituents, but cannot impose groupings that cross-cut the durational tree. No sequence of rewrites can produce a partition of the frontier that does not correspond to some subtree. Any formalism that operates entirely within

the subdivision hierarchy will face the same constraint. When musical practice demands groupings that cut across subdivision boundaries, as the Carnatic techniques in Section 3 systematically do, a second structural layer is required.

Rhythm trees, as developed in the computer-aided composition literature [7, 6], provide a structural representation of rhythmic notation without commitment to models of rhythmic cognition. The extensions proposed here maintain this agnosticism: the accent tree is a generative device for imposing groupings that cross-cut the subdivision hierarchy, not a model of metrical perception or structural hearing. In particular, we make no claim about the relationship between accent-tree structure and felt or heard metrical layers. Whether and how these constructions might support analytical or cognitive accounts of layered metric hierarchy remains an open question to which we return in the conclusion.

3. CARNATIC RHYTHMIC ORGANIZATION

We draw our test cases from Carnatic rhythm, focusing on Reina’s systematic treatment [8, 9]. Carnatic music provides a particularly stringent test case for rhythmic formalisms because its theoretical vocabulary already distinguishes durational and accentual organization as independent parameters and because its compositional techniques explicitly manipulate the tension between them.

Rhythm in Carnatic practice is organized around the *tala*, a cyclical metric framework comprising a fixed number of beats.² The first beat of the cycle, called *sam*, serves as the primary point of resolution; phrases are structured to align with *sam*. Each beat in the *tala* is subdivided into equal pulse units according to a *gati*: *tisra* (3), *chatusra* (4), *khanda* (5), *misra* (7), or *sankirna* (9). These subdivisions define the *matra* level, the finest regular grid on which rhythmic patterns are realized. A separate parameter, the *jathi*, governs how matras are grouped by accent. While *gati* determines subdivision *within* beats, *jathi* determines how those subdivisions are *grouped*. The two parameters are independent: the same 20-matra grid can receive accent groupings of (4, 4, 4, 4, 4) (aligned with beats), (5, 5, 5, 5) (cross-cutting beats), or (1, 7, 3, 4, 3, 2) (irregular).

This independence is the crux of the formal problem. A rhythm tree encodes the cycle, beat, and matra hierarchy naturally: the root represents the cycle, depth-1 nodes represent beats, and depth-2 nodes represent matras. But an accent pattern like (1, 7, 3, 4, 3, 2) partitions the 20 matras into spans that correspond to no subtree. The technique of *jathi bhedam* exploits exactly this: irregular accent sequences are superimposed on a uniform *gati* grid, creating the impression of continuous meter change while the underlying pulse remains fixed. Each accent span is performed with “downbeat feeling” [8, 9], yet the spans cross-cut beat boundaries. The rhythm tree stays fixed; only the grouping varies.

² This account is necessarily simplified. In practice, *tala* also encodes structural constraints that govern how and when surface-level rhythmic detail may develop. Since these issues are not central to the present discussion, we set them aside here.

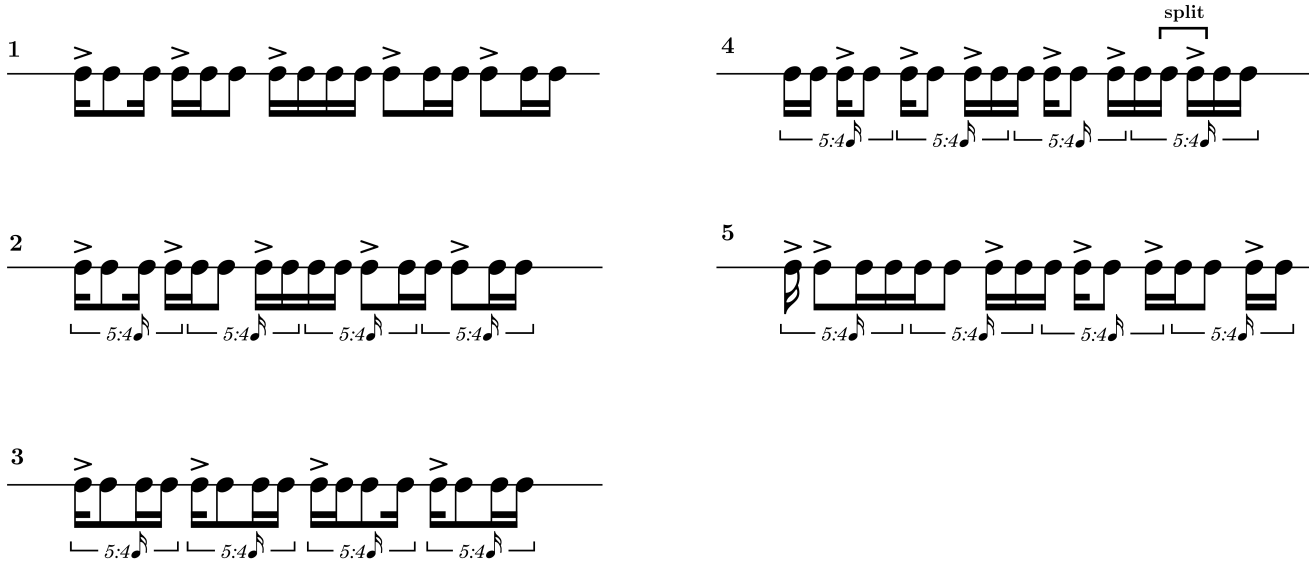


Figure 1. Development of a rhythmic motif through rhythmic sangati. (The beaming follows the convention used by Reina for Carnatic music in Western notation [8, 9].) (1) The original motif in chatusra (four subdivisions per beat), with accents every four matras. (2) The motif rephrased in khanda (five subdivisions per beat) while preserving the accent every four matras; the original gati becomes the jathi, creating the effect of a local metric modulation. (3) The motif takes on the accent pattern of the new gati, with accents now falling every five matras, leading to motivic development. (4) Both gati and jathi are changed simultaneously. When the original phrase does not fit evenly within the new grouping, notes may be split (as marked) to align with the new jathi accent pattern, with any remainder phrased as an anacrusis to ensure the sangati coincides with tala sam. (5) The motif is rephrased as jathi bhedaṁ, applying an irregular accent pattern that does not coincide with the khanda gati.

What makes Carnatic rhythm particularly revealing for our purposes is that these accent spans are not merely perceptual effects; they are first-class compositional objects. Spans can be reordered, fragmented, and treated as units of subdividable time, and a single rhythmic phrase can be re-phrased across multiple gati environments. This is the technique of rhythmic sangati: a phrase constructed in one gati is adapted to another, with the relationship between original and target accent structures varying according to compositional intent. In Figure 1, a phrase in chatusra (four subdivisions per beat) becomes a phrase in khanda (five subdivisions per beat). The phrase spans different numbers of beats in each gati, but it is recognized as the “same rhythm”—similar in feeling to metric modulation in Western music.

Reina distinguishes four approaches to such transformations: 1, the original gati becomes the jathi in the new context, preserving the accent pattern exactly but at a different tempo relationship; 2, the note-value sequence is preserved but rebeamed into the new gati’s grouping, imposing a fresh accent structure; 3, both gati and jathi change, requiring reconciliation when the new grouping does not evenly divide the phrase length; 4, the source phrase is itself constructed using jathi bhedaṁ, and its irregular accent structure is preserved through the gati transformation, yielding jathi bhedaṁ phrasing in the target gati. Each approach implies a different formal relationship between source and target accent structures, from identity, to replacement, to constrained compatibility, to the preservation of already-irregular groupings across metric contexts.

Rhythmic sangati poses a precise combinatorial question: given two accent structures over the same number of pulses, which onset patterns are compatible with both? In Section 4 we introduce *accent trees*, which formalize this interplay between rhythmic content and accentual grouping. The model addresses sangati and related techniques in two complementary ways: First, an *attachment model* allows rhythmic content to be time-scaled into accent spans of varying sizes, capturing the tempo-fluctuation effects that arise when the same phrase is stretched or compressed across different gati environments. This also provides a natural account of gati bhedaṁ: a tala cycle can be progressively elaborated by attaching rhythm trees into accent spans, increasing rhythmic depth while the accent skeleton remains fixed. Second, we observe that accent compatibility reduces to a simple combinatorial condition: a rhythm is compatible with both accent structures exactly when its onset positions contain all required accent heads. We encode this as a bitmask and enumerate the compatible rhythms by iterating over subsets of the free positions. Together, these mechanisms yield not a single sangati but a structured compositional space: the complete set of rhythms satisfying the sangati constraint, each available for attachment into the accent spans of either context.

4. ACCENT TREES

We work with two distinct structural layers over rhythmic time: a durational layer, represented by rhythm trees, and an accent layer, represented by accent trees. The durational

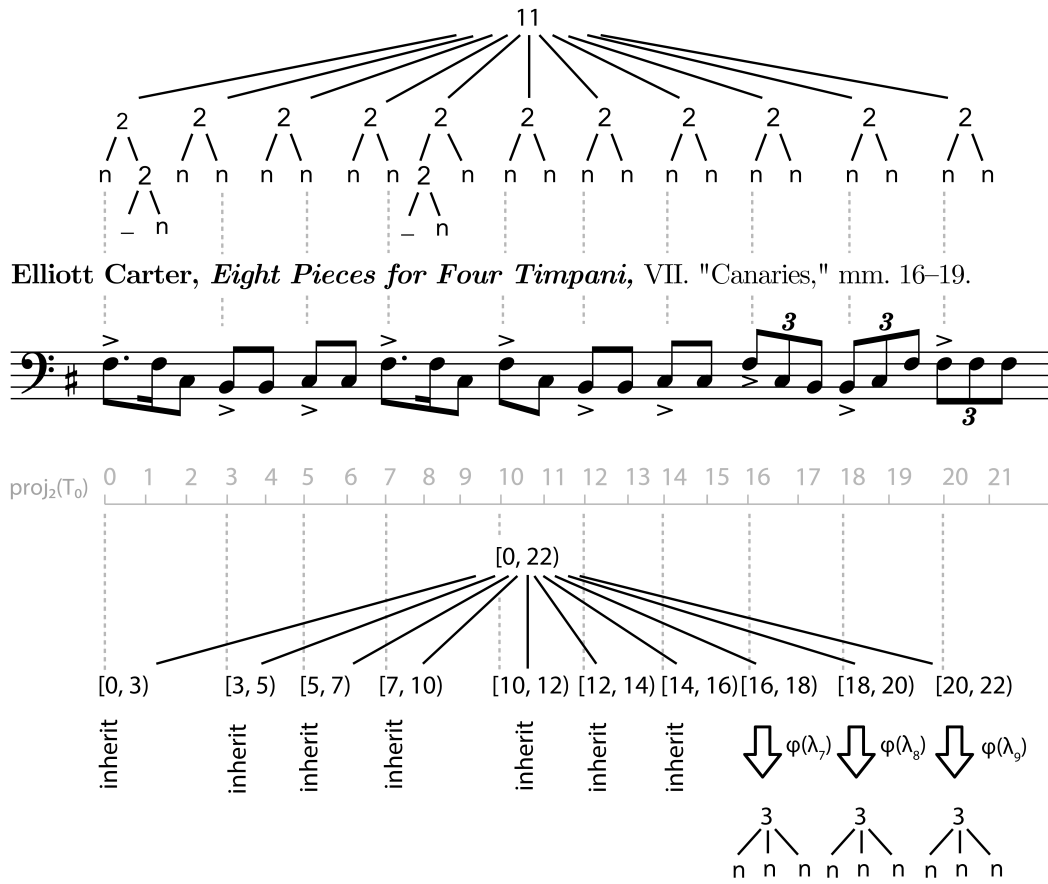


Figure 2. Elliott Carter, *Eight Pieces for Four Timpani* [10], VII. “Canaries,” mm. 16–19, reconstructed as a single-level accent product $A \triangleright T_0$ at depth 2. Top: the base rhythm tree $T_0 = 11(2(n, n), \dots, 2(n, n))$, with dotted figures at depth 3 in beats 1 and 4. Middle: the score, with positions $\nu_0, \dots, \nu_{21}$ of $\text{proj}_2(T_0)$ indexed below the staff. Bottom: the accent tree, whose root $[0, 22]$ partitions the projection into spans $(3, 2, 2, 3, 2, 2, 2, 2, 2, 2)$. Where spans cross-cut beat boundaries at depth 1, no span corresponds to any subtree of T_0 . The projection presents each position as an undifferentiated index; rhythmic detail below depth 2 is invisible to the accent tree and passes through inheritance unchanged via (22). Spans λ_7 – λ_9 replace inherited content with $\phi(\lambda) = 3(n, n, n)$, producing triplets within the unchanged base grid. Figure 3 extends this construction to a later passage in the same movement, where recursive application of the accent tree captures a metric modulation from tisra to chatusra jathi 5.

Elliott Carter, *Eight Pieces for Four Timpani*, VII. "Canaries," mm. 43, 45–48.

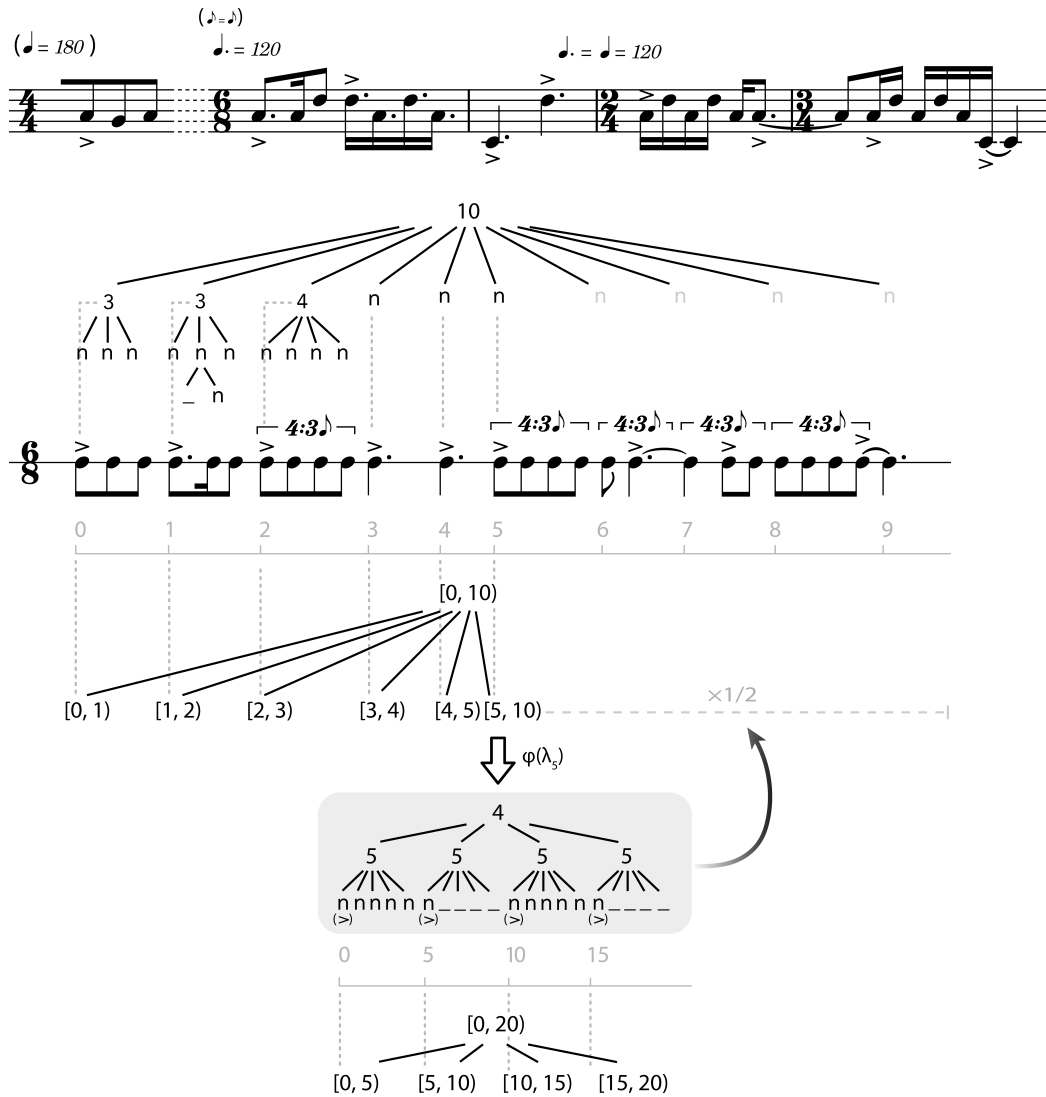


Figure 3. Recursive accent-tree reconstruction of Elliott Carter, *Eight Pieces for Four Timpani* [10], VII. "Canaries," mm. 43, 45–48. Read through a Carnatic lens, the passage begins in *tisra gati* at second speed. A single beat of *chatusra* subdivision appears at the second beat following the 6/8 bar line, after which two unsubdivided beats introduce metric ambiguity. The metric modulation $\text{♩} = \text{♩}$ resolves this ambiguity into *chatusra jathi* 5, a regrouping of the pulse into spans of five within a fourfold subdivision. The reconstruction (middle) normalizes the passage into a single 6/8 span of ten beats. The accent tree $[0, 10)$ imposes an asymmetric partition whose span $[5, 10)$ receives an attached rhythm tree $\varphi(\lambda_5)$ comprising four groups of five matras, time-scaled by $\times \frac{1}{2}$. A second accent tree $[0, 20)$ partitions the frontier of $\varphi(\lambda_5)$ into four equal spans (bottom). This example illustrates a broader point about the data structure: because the accent tree with its attachment function can represent asymmetric partitioning, recursive elaboration, and time-scaling within a single hierarchical object, metric modulation need not be treated as an external event that interrupts one metric regime and installs another. Instead, it arises internally as a consequence of how accent spans at successive levels relate to one another.

layer specifies how a span of time is recursively subdivided into notated units (notes and ties). The accent layer specifies how those units are grouped into spans that carry downbeat feeling or other forms of rhythmic salience. This mirrors the syntax/prosody split in speech: a syntactic tree organizes words, while a prosodic tree groups them into higher-level units such as prosodic words and phrases. Here the rhythm tree plays the role of syntax, and the accent tree plays the role of prosody.

An attachment model interprets an accent tree over a base rhythm tree by operating on labeled duration sequences (2). Figure 2 illustrates a single-level accent product in which an accent tree partitions a base rhythm tree’s depth-2 projection into spans that cross-cut beat boundaries, with some spans inheriting the base rhythm and others receiving attached content. The output is a new labeled duration sequence, which can then be mapped back to a notational rhythm tree by a reconstruction procedure in the style of Ycart et al. [11].

4.1 Depth Projection and Repleteness

Analytical studies of West African drum ensemble music commonly assume a fastest pulse—variously called the “density referent” or “equal-pulse base”—that serves as an orientation grid from which rhythms are composed, independent of surface articulation [12]. Carnatic rhythmic theory makes a similar commitment: accent groupings are defined over matras, the pulse positions of the subdivision grid, not over note onsets.

For this reason, the accent layer operates over a projection of the rhythm tree at a chosen depth, rather than directly over frontier positions. This design reflects a key feature of Carnatic rhythmic theory: accent groupings are defined over matras, the pulse positions of the subdivision grid, not over note onsets. A jathi pattern (3, 4, 5, 3, 5) partitions twenty matras into five spans. These spans are counted in matras regardless of how notes articulate within them. Of course, for an accent to be audible, the first matra of each span must carry an onset. This is why our attachment model requires attached rhythms to begin with an onset (Section 4.2), and why the sangati derivation in Section 4.5 constrains onset positions to include all accent heads. The grouping structure itself, however, is defined at the matra level, independent of surface articulation.

This is why we require the base rhythm tree to be *replete* at the projection depth. Repleteness ensures that every matra exists as an addressable node in the tree, so that accent spans can be measured correctly even when the musical surface contains long notes spanning multiple pulses. Without repleteness, a half-note occupying four matras would appear as a single leaf, and the accent tree would have no way to count those four positions toward span length.

Recall that the depth of a node is its distance from the root, with the root at depth 0. A rhythm tree T is *replete at depth d* if every leaf has depth at least d . Any rhythm tree can be made replete by applying the rewrite rules of Jacquemard et al. [6]. A leaf at insufficient depth is replaced by a subtree of tied notes that preserves the original duration while exposing each matra as a separate node

(Figure 4). Since the rewrite system is sound, this transformation preserves rhythmic equivalence. The sounding result is unchanged, but the tree now supports matra-level accent assignment.

Let T be a rhythm tree replete at depth d . The *d -projection* of T is the sequence

$$\text{proj}_d(T) = (\nu_0, \dots, \nu_{k-1}) \quad (3)$$

of nodes at depth exactly d , collected in left-to-right order. Each node ν_i roots a subtree $T[\nu_i]$ covering a contiguous block of the frontier. The blocks for $i = 0, \dots, k-1$ partition $\text{fr}(T)$.

We identify the projection with its *index set*

$$I_d(T) = \{0, 1, \dots, k-1\}, \quad (4)$$

where index i refers to the subtree $T[\nu_i]$.

Different choices of d pick out different levels of metrical structure: a full tala cycle ($d = 0$), beats within a tala ($d = 1$), subdivision pulses ($d = 2$), and so on.

We associate the depth- d projection (3) with a *projected duration sequence*, which records the durational content carried by each projected node. Specifically, the *d -projected duration sequence* of T is

$$\text{pds}_d(T) = ((T[\nu_0], D_0), \dots, (T[\nu_{k-1}], D_{k-1})), \quad (5)$$

where $T[\nu_i]$ is the subtree rooted at ν_i and D_i is the total duration of its frontier.

Each pair $(T[\nu_i], D_i)$ is a metrically meaningful unit at the chosen level. Together the projected subtrees cover the entire cycle.

4.2 The Attachment Model

Accent trees partition the index set $I_d(T_0)$ (4) of projection positions into grouped spans. Each node of the accent tree represents one span, and $\text{slot}(\alpha)$ records its length (in number of projection positions).

We write $A \triangleright T_0$ to denote the accent action of A on T_0 and call the result an *accent product*. The action is mediated by the depth projection: A partitions $I_d(T_0)$, and the attachment function ϕ specifies the content of each span.

Formally, an accent tree at depth d for a rhythm tree T_0 is a rooted, ordered tree A with a function

$$\text{slot} : \text{nodes}(A) \rightarrow \{1, 2, \dots\}. \quad (6)$$

We require:

1. **Root span.** The root covers the entire projection:

$$\text{slot}(\text{root}(A)) = k. \quad (7)$$

2. **Tiling.** If an internal node α has ordered children $\beta_1, \dots, \beta_q$, then the children tile (partition) α ’s span:

$$\sum_{i=1}^q \text{slot}(\beta_i) = \text{slot}(\alpha). \quad (8)$$

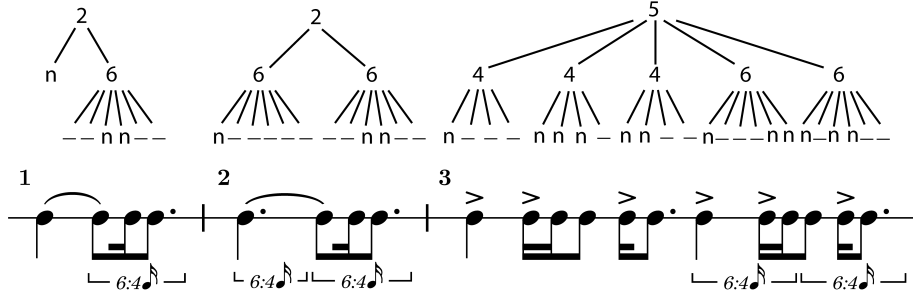


Figure 4. Repleteness at depth 2. (1) The tree $2(n, 6(_, _, n, n, _, _))$ is non-replete at depth 2. The first child has a terminal at depth 1, so the matras it spans at depth 2 are not individually addressable. (2) The same rhythm rewritten as a tree replete at depth 2: $2(6(n, _, _, _, _, _), 6(n, _, n, n, _, _))$. The first note is now represented as equal subdivisions at depth 2, exposing each matra as a separate node. Both trees produce the same labeled duration sequence; only the internal structure differs. (3) A rhythmic sangati: the accent cycle (4, 4, 4) partitions twelve matras while internal articulation varies across beats. Repleteness makes this partition well-defined. Groupings are counted in matras, even though audible accents require onsets.

Condition (2) is the *tiling property*: children’s slots partition exactly the parent’s slot, in left-to-right order. Thus the leaves of A induce a partition of $I_d(T_0)$ into contiguous blocks.

The only constraint on this partition is that the slot values sum to k ; individual spans may be of any length from 1 to k . Since an attached tree filling a span is itself a rhythm tree, it admits its own depth projection and can serve as the base for a further accent product (Section 4.4). Arbitrary span length is thus the mechanism that enables recursive elaboration: a large span can host a complex attached tree, which can in turn be repartitioned by its own accent tree, to any depth.

Since the base tree is replete at depth d , each projected position ν_i has duration D_i . When all projected positions have equal duration (for example, at the beat level, or at the matra level under uniform gati), this simplifies to

$$D_{\text{proj}} = \frac{1}{k}, \quad (9)$$

and the span duration of any accent node α is

$$D_\alpha = \text{slot}(\alpha) D_{\text{proj}} = \frac{\text{slot}(\alpha)}{k}. \quad (10)$$

When gati varies across beats, matra durations differ and the general form

$$D_\lambda = \sum_{i=a}^{b-1} D_i \quad (11)$$

must be used for a leaf λ with global interval $[a, b)$.

Let α be an internal node with children $\beta_1, \dots, \beta_q$. The local interval of β_i relative to α is

$$\text{loc}(\beta_i \mid \alpha) = [s_i, s_i + \text{slot}(\beta_i)), \quad (12)$$

where $s_1 = 0$ and

$$s_{i+1} = s_i + \text{slot}(\beta_i). \quad (13)$$

Global intervals are defined compositionally: the root has interval $[0, k)$. If α has global interval $[a, b)$ and β_i has local interval $[s_i, s_i + \text{slot}(\beta_i))$, then

$$\text{glob}(\beta_i) = [a + s_i, a + s_i + \text{slot}(\beta_i)). \quad (14)$$

The global interval of a leaf λ specifies exactly which indices of $I_d(T_0)$ —hence which projected subtrees—fall under that accent span.

The accent-tree formalism defines grouping at a chosen projection depth, typically the beat level, while the rhythmic content that fills each span is measured at a finer level, typically the matra. This separation requires reasoning simultaneously about two quantities: the *slot*, which counts how many projection-level units a span covers, and the *capacity*, which counts how many matras that span contains, given the gati of each covered beat.

When the accent tree operates at depth d , the number of matras within a span depends on the arity of each covered beat. For a leaf λ with global interval $[a, b)$, we define the capacity as

$$\text{cap}(\lambda) = \sum_{i=a}^{b-1} \text{arity}(\nu_i), \quad (15)$$

where ν_i is the i th node in the depth- d projection and $\text{arity}(\nu_i)$ is the subdivision count of the corresponding beat.

This two-level bookkeeping is a direct consequence of the attachment model. An attached rhythm tree is time-scaled to fill a span whose duration is determined by slot. Two spans with the same slot duration but different capacities—for example, five tisra beats versus five khanda beats—have the same absolute duration but different numbers of matras, so the same attached rhythm will align differently with the underlying pulse grid.

To connect these abstract accent positions to concrete surface events, we define an anchoring map from each leaf-span to the realized leaf event that carries its accent head. Fix an accent attachment model (T_0, d, A, ϕ) . For each leaf $\lambda \in \text{leaves}(A)$ with global interval

$$\text{glob}(\lambda) = [a, b), \quad (16)$$

define its anchored accent head $\text{head}(\lambda)$ as the leftmost realized leaf event contributed by λ . If $\phi(\lambda)$ is undefined (inheritance), then $\text{head}(\lambda)$ is the first leaf of $\text{lds}(T_0[\nu_a])$. If $\phi(\lambda) = t$ (attachment), then t is time-scaled to fill the span duration D_λ : each leaf of t retains its proportional

share of duration, but the absolute durations are multiplied by D_λ .

An accent tree specifies *where* accent spans fall; an *attachment function* specifies *what* fills them. Formally, let

$$\mathcal{R} = \{t \in \text{RhythmTree} : \ell_0(t) = n\} \quad (17)$$

be the set of rhythm trees whose leftmost leaf is an onset. An *attachment function* for an accent tree A is a partial map

$$\phi : \text{leaves}(A) \rightarrow \mathcal{R}. \quad (18)$$

Requiring attached rhythms to begin with an onset ensures that each accent span carries an audible attack at its head, which is the defining property of an accented position.

An *accent attachment model* is a tuple (T_0, d, A, ϕ) consisting of a base rhythm tree T_0 replete at depth d , an accent tree A at that depth, and an attachment function ϕ . For each leaf $\lambda \in \text{leaves}(A)$, let

$$[a, b) = \text{glob}(\lambda) \quad (19)$$

be its global interval, so that λ claims projection positions

$$a, a + 1, \dots, b - 1. \quad (20)$$

The span duration of λ is

$$D_\lambda = \sum_{i=a}^{b-1} D_i, \quad (21)$$

where D_i is the duration of the i th projected subtree.

For a leaf λ with global interval $[a, b)$, we define its realized sequence σ_λ by distinguishing two cases, depending on whether λ inherits material from the base tree or receives an attached rhythm.

If $\phi(\lambda)$ is undefined (inheritance), then σ_λ is the concatenation of the labeled duration sequences (2) of the projected subtrees covered by λ :

$$\sigma_\lambda = \text{lds}(T_0[\nu_a]) \cdot \text{lds}(T_0[\nu_{a+1}]) \cdots \text{lds}(T_0[\nu_{b-1}]). \quad (22)$$

If $\phi(\lambda) = t$ (attachment), then t is time-scaled to fill the span of duration D_λ .

Inheritance preserves the durations already present in the base tree, while attachment rescales the attached rhythm to exactly fill the accent span.

These two cases—inheritance and attachment—correspond to distinct compositional operations. Inheritance regroups existing material under a new accent structure; attachment fills accent spans with new content, time-scaled to fit. Carnatic practice provides particularly clear examples of both.

Inheritance is the default case: the accent tree partitions existing rhythmic material without replacing it. The durations in the base tree remain unchanged; only the grouping varies. A phrase can be partitioned into accent spans of any length—aligned with the subdivision grid or cross-cutting it—without altering its durational content. The accent tree imposes a grouping structure; inheritance copies the underlying rhythm into each span.

Attachment handles cases where the accent span itself becomes a frame for subdivision—where new rhythmic content is inserted and time-scaled to fill the span. Attachment

is not mere concatenation. Sequential rhythm trees are independent durational structures joined end-to-end; an attached rhythm is scaled to fill a span whose boundaries are defined by a separate accent hierarchy. In a single-layer model, rhythmic structure is purely durational: grouping is a byproduct of subdivision. The two-layer model treats accent structure as a parallel hierarchy with its own compositional logic. Accent spans can be subdivided, reordered, nested, and transformed independently of the rhythms that fill them. But the two layers share an interface: accent heads must be articulable, so they must coincide with onsets. This coupling means that transformations in either layer propagate constraints to the other. Elaborating an accent structure may force rhythmic adaptation; imposing a new rhythm may render certain accent patterns unsupportable.

Carnatic practice foregrounds this interaction. In the technique of *mixed jathi nadai bhedom* [8, 9], two accent groupings alternate to tile a cycle exactly—say, spans of 3 and 7 matras repeating twice to fill a span of 20 matras. Each span can then be subdivided independently: the 7-span might be split into three, while the 3-span is split into four, yielding the impression of two alternating tempos (Figure 5). The accent spans are first-class compositional units, subject to motivic development through subdivision, augmentation, and transformation. The rhythms that fill them must articulate each span head, but are otherwise free to vary. This is the long-distance dependency that context-free grammars cannot express: the accent structure imposes constraints across the entire cycle, and any surface rhythm must satisfy all of them simultaneously. The attachment model represents this directly—each span hosts its own attached rhythm tree, time-scaled to fit, while the accent skeleton remains fixed.

The accent tree separates *where* accent spans fall from *what* fills them. Inheritance and attachment are the two ways content can relate to that scaffold: either the existing rhythmic surface is regrouped (*jathi bhedom*), or new content is inserted and scaled to fit. Both operations preserve the matra grid; both produce labeled duration sequences that compile back to rhythm trees.

When subdivision varies across spans, as in *mixed jathi nadai bhedom*, the accent tree becomes the structural bridge between subdivision contexts that share a temporal framework but not an internal pulse. Carnatic practice pushes this independence to its perceptual limit: adjacent spans in *mixed jathi nadai bhedom* are performed at different effective tempos, yet the tala cycle and its beat grid remain fixed. The effect approaches polymeter, but the structural commitment is to a shared grid with accentual repartitioning on top. The two-layer model captures this: the attachment function imposes no constraint that attached trees share a common arity, but the projection that anchors the accent tree remains a well-defined level of the base tree. Whether other repertoires require the stronger reading, namely the accent skeleton as the sole structural link between incommensurable subdivisions, is a question beyond the scope of this paper.

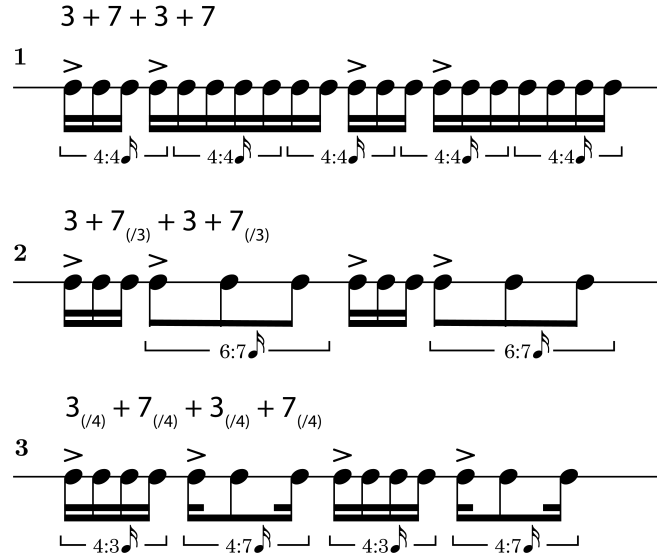


Figure 5. *Mixed jathi nadai bhedom* is a Carnatic technique in which two accent groupings alternate to tile a cycle exactly, with each span subject to independent subdivision. The technique illustrates how accent spans function as first-class compositional units under the attachment model. Here, spans of 3 and 7 matras alternate: (1) the basic alternation; (2) the 7-span subdivided into three; (3) both spans subdivided into four. The result is something perceptually like metric modulation, yet the accent skeleton remains fixed throughout—the accent tree defines the span boundaries, and each span hosts an independently elaborated rhythm, time-scaled to fit.

4.3 Compilation and Reconstruction

The accent attachment model associates a realized sequence σ_λ to each leaf of the accent tree. Let $A \triangleright T_0$ be an accent product, and let $\lambda_1, \dots, \lambda_m$ be the leaves of A in left-to-right order. The compiled output is the concatenation

$$\sigma_{\text{out}} = \sigma_{\lambda_1} \cdot \sigma_{\lambda_2} \cdots \sigma_{\lambda_m}. \quad (23)$$

By the tiling property (8), the leaves’ global intervals partition $[0, k)$, so

$$\text{dur}(\sigma_{\text{out}}) = \text{dur}(\text{lds}(T_0)), \quad (24)$$

that is, attachments redistribute events within each span but preserve total duration. The accent layer thus specifies *what* to play and *when*; it remains to find a rhythm tree, a notational object, that realizes this sequence.

The compiled output σ_{out} is a labeled duration sequence; to obtain conventional notation, one must find a rhythm tree t with $\text{lds}(t) = \sigma_{\text{out}}$. We adopt the k -best enumeration algorithm of Ycart et al. [11], which uses dynamic programming over a subdivision schema to produce candidate trees in order of increasing notational complexity. Since σ_{out} already specifies exact onset times, the algorithm reduces to selecting the simplest tree consistent with the compiled sequence.

4.4 Recursive Accent Products

Since an attached rhythm tree is itself a rhythm tree, it can serve as the base for its own accent product. An attachment $\varphi^0(\lambda) = T_1$ may host a further action $A^1 \triangleright T_1$, elaborating the content within λ ’s span. This nesting requires no additional machinery: the same attachment model applies at

each level, compiling bottom-up. Complex rhythmic structures emerge as a byproduct of recursive attachment rather than as special cases requiring dedicated treatment. Figure 3 illustrates this with mm. 43, 45–48 of Carter’s “Canaries” [10], where a passage conventionally described as metric modulation—or, in Carnatic terms, a shift from tisra gati to chatusra jathi 5—arises naturally from nested accent trees at successive levels of attachment.

A second mode, *iterative recursion*, compiles $A^0 \triangleright T_0$ to a rhythm tree T'_0 and applies a new action $A^1 \triangleright T'_0$ to the result. Since compilation flattens span boundaries into a labeled duration sequence, A^1 may impose groupings that cross-cut the spans induced by A^0 .

4.5 Derivation of Rhythmic Sangati from Figure 6

The attachment model provides one account of sangati: a phrase constructed in one gati can be attached into accent spans defined by another, with time-scaling handling the tempo relationship. But sangatis also pose a complementary question: given two accent structures over the same span, which onset patterns are compatible with both? This is simply a matter of finding rhythms whose onsets include all required accent heads from both structures. The formalism reduces this to a simple combinatorial problem, and the solution defines a structured compositional space rather than a single rhythm.

Figure 6 shows a single tala cycle partitioned into two gati regions: five beats in tisra (three pulses per beat) and three beats in khanda (five pulses per beat). Through attachment, each span takes on a capacity of fifteen matras ($5 \times 3 = 3 \times 5 = 15$), but organizes them differently. One way to construct a rhythmic sangati is to find an onset pattern that works in both metric environments, aligning with beat

T_0 replete at $d=1$

$$T_0 = 8(\underbrace{3(\dots), 3(\dots), 3(\dots), 3(\dots), 3(\dots)}_{\lambda_1 \text{ tistra}}, \underbrace{5(\dots), 5(\dots), 5(\dots)}_{\lambda_2 \text{ khanda}})$$

$$\text{slot}(\lambda_1) = 5, \quad \text{glob}(\lambda_1) = [0, 5)$$

$$\text{slot}(\lambda_2) = 3, \quad \text{glob}(\lambda_2) = [5, 8)$$

Attachment

$$\phi(\lambda_1) = T_1 = 5(3(n, n, n), 3(n, n, n), 3(n, n, n), 3(n, n, n), 3(n, n, n))$$

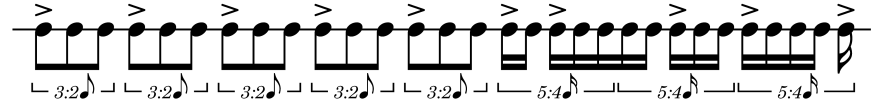
$$\phi(\lambda_2) = T_2 = 3(5(n, n, n, n, n) \ 5(n, n, n, n, n) \ 5(n, n, n, n, n))$$

$$\text{cap}(\lambda_1) = 5 \times 3 = 15 = 3 \times 5 = \text{cap}(\lambda_2)$$

Recursive accent trees

$$A^1 \text{ on } \lambda_1: \ 3+3+3+3+3 \implies \text{ons}(A^1) = \{0, 3, 6, 9, 12\}$$

$$A^2 \text{ on } \lambda_2: \ 2+5+3+4+1 \implies \text{ons}(A^2) = \{0, 2, 7, 10, 14\}$$



Union of accent masks

$$U = \text{ons}(A^{\text{tistra}}) \cup \text{ons}(A^{\text{khanda}}) = \{0, 2, 3, 6, 7, 9, 10, 12, 14\}$$

Examples of compatible onset patterns

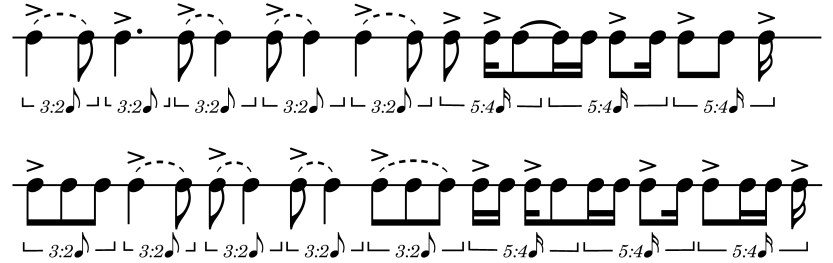


Figure 6. Derivation of a sangati-compatible rhythm space. An accent tree at depth 1 partitions eight beats into spans of five (tistra, slot = 5) and three (khanda, slot = 3). Through attachment, λ_1 receives five triplets and λ_2 three quintuplets, giving both spans equal capacity: $5 \times 3 = 3 \times 5 = 15$ matras. Recursive accent trees then act on the attached trees: regular gati accents every three matras on T_1 , yielding onset set $\{0, 3, 6, 9, 12\}$, and an irregular jathi (2, 5, 3, 4, 1) on T_2 , yielding $\{0, 2, 7, 10, 14\}$. Their union determines nine required accent positions; the six free positions yield $2^6 = 64$ compatible rhythms. Two examples are shown.

Added onsets	$(\ell_0, \ell_1, \dots, \ell_{14})$	total
$\emptyset$	$(n, -, n, n, -, -, n, n, -, n, n, -, n, -, n)$	9
$\{1, 4\}$	$(n, n, n, n, n, -, n, n, -, n, n, -, n, -, n)$	11
$\{4, 8, 11, 13\}$	$(n, -, n, n, n, n, -, n, n, n, n, n, n, n, n)$	13
F	$(n, n, n, n, n, n, n, n, n, n, n, n, n, n)$	15

Table 1. Four rhythms from the 64-pattern sangati space. All share the nine required onsets; they differ only in which free positions receive additional onsets.

accents in the first half and with the irregular jathi pattern (2, 5, 3, 4, 1) in the second.

We represent onset patterns as binary words of length n , where bit i is 1 if position i carries an onset and 0 if it carries a tie. The accent trees for the two halves induce onset sets:

$$\begin{aligned} \text{ons}(A_{\text{tisra}}) &= \{0, 3, 6, 9, 12\}, \\ \text{ons}(A_{\text{khandā}}) &= \{0, 2, 7, 10, 14\}. \end{aligned} \quad (25)$$

A rhythm compatible with both must have onsets at every position in their union:

$$U = \{0, 2, 3, 6, 7, 9, 10, 12, 14\}. \quad (26)$$

We encode U as a bitmask M by setting bit i to 1 for each $i \in U$. Bits are indexed left-to-right by pulse position: bit i corresponds to position i , with bit 0 the leftmost bit.

$$M = (101100110110101)_2. \quad (27)$$

The positions where M has bit 0 are *free*, meaning each may independently carry an onset or a tie:

$$F = \{1, 4, 5, 8, 11, 13\}. \quad (28)$$

This yields $2^{|F|} = 64$ compatible rhythms that fit both accent spaces.

Enumerating this space is straightforward. We iterate a $|F|$ -bit counter s from 0 to $2^{|F|} - 1$. For each s , we construct a rhythm by starting with the required mask M and inserting additional onsets at free positions according to the bits of s . The key operation is a bitwise OR: we shift a 1 into each free position selected by s and combine the result with M .

Table 1 shows four rhythms from this space. The sparsest (adding no free positions, $\emptyset$) places onsets only at required positions, while the densest (adding all free positions, F) fills every pulse. The 62 intermediate patterns offer a graduated range of rhythmic density, providing composers with a principled vocabulary of sangati-compatible phrases.

The enumeration runs in $O(n \cdot 2^{|F|})$ time, which is output-optimal. In Carnatic contexts this is tractable: projection lengths rarely exceed 20 to 30 matras, each accent structure places 5 to 10 heads, and their union typically covers half or more of the available positions, leaving $|F| \leq 10$ to 15 free positions. The musically interesting cases, where accent patterns create tension by overlapping only partially, are precisely those that shrink the solution space further.

When $|F|$ is larger, or when only a subset of rhythms is desired, the enumeration admits pruning. The iteration

over subsets of F forms a binary decision tree; constraints on tie density or maximum run length can prune branches early. Alternatively, a best-first search ordered by a scoring function, such as proximity to a target density or resemblance to a seed pattern, could yield preferred rhythms without exhaustive enumeration.

5. CONCLUSION

A useful test of a formalism is the range of structural phenomena it can make explicit. Carnatic rhythmic theory supplies a concentrated set of such cases, but the formal problem is not Carnatic-specific. Reina [8] demonstrates a broader analytical reach for Carnatic rhythmic structures, applying it systematically across Western repertoire: jathi bhedam in Messiaen, Stravinsky, Ligeti, and Maxwell Davies; independent gati layering in Carter and Donatoni; nadai bhedam and anuloma-pratiloma structures in Xenakis, Castiglione, and Ferneyhough. In each case, the rhythmic interest lies not in the durational structure or the accent structure alone, but in the tension between them—a tension that no single-tree formalism over the subdivision hierarchy can structurally represent. The duration–accent interface proposed here provides a compact representation for these phenomena while remaining compatible with existing rhythm-tree grammars and their associated rewrite and enumeration algorithms.

Future work falls into two main directions. First, accent trees themselves can be placed under an explicit generative grammar, for example by adapting Tree-Adjoining Grammar to treat accent cycles as elementary trees and to model recursive elaboration, nesting, and span-level transformations across metrical contexts. Second, the present model suggests quantitative alignment constraints that measure systematic match or mismatch between accent boundaries and durational constituents at multiple depths, yielding a computable notion of rhythmic tension analogous to prosodic markedness. Wells [13, 14] develops a complementary intervallic approach to Carnatic rhythm, using Lewin’s generalized interval systems to model the trajectory of melodic phrases through the tala cycle during trikala performance. While his Met intervals track dynamic displacement—how phrase boundaries migrate across beat positions under speed transformation—accent trees capture the hierarchical grouping structure that listeners recognize as preserved across such transformations. Integrating these perspectives could yield a richer account of sangati, combining Wells’s precise tracking of metric tension with the structural characterization proposed here.

6. REFERENCES

- [1] M. Yeston, *The Stratification of Musical Rhythm*. New Haven: Yale University Press, 1976.
- [2] M. Rohrmeier, “Towards a formalization of musical rhythm,” in *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*, Montréal, Canada, 2020.

- [3] E. O. Selkirk, “The syntax-phonology interface,” in *The Handbook of Phonological Theory*, 2nd ed., J. Goldsmith, J. Riggle, and A. C. L. Yu, Eds. Oxford: Blackwell, 2011.
- [4] J. Tepperman and S. Narayanan, “Tree grammars as models of prosodic structure,” in *Proceedings of Interspeech*, 2008.
- [5] F. Jacquemard, A. Ycart, and M. Sakai, “Generating equivalent rhythmic notations based on rhythm tree languages,” in *Proceedings of the Third International Conference on Technologies for Music Notation and Representation (TENOR’17)*, A Coruña, Spain, 2017, pp. 145–153.
- [6] F. Jacquemard, P. Donat-Bouillud, and J. Bresson, “A structural theory of rhythm notation based on tree representations and term rewriting,” in *Mathematics and Computation in Music: 5th International Conference, MCM 2015, London, UK, June 22–25, 2015, Proceedings*. Springer, 2015.
- [7] C. Agon, K. Haddad, and G. Assayag, “Representation and rendering of rhythmic structures,” in *Proceedings of the Second International Conference on Web Delivering of Music (WEDELMUSIC)*. Darmstadt, Germany: IEEE Computer Press, 2002, pp. 109–113.
- [8] R. Reina, “Karnatic rhythmical structures as a source for new thinking in western music,” PhD thesis, Middlesex University, 2016.
- [9] —, *Applying Karnatic Rhythmical Techniques to Western Music*. New York: Routledge, 2016.
- [10] E. Carter, *Eight Pieces for Four Timpani*. New York: Associated Music Publishers, 1968.
- [11] A. Ycart, F. Jacquemard, J. Bresson, and S. Staworko, “A supervised approach for rhythm transcription based on tree series enumeration,” in *Proc. Sound and Music Computing Conference (SMC)*, Hamburg, Germany, 2016.
- [12] R. Polak, “Rhythmic feel as meter: Non-isochronous beat subdivision in jembe music from Mali,” *Music Theory Online*, vol. 16, no. 4, 2010. [Online]. Available: <https://www.mtosmt.org/issues/mto.10.16.4/mto.10.16.4.polak.html>
- [13] R. L. Wells, “A generalized intervallic approach to metric conflict in Liszt,” *Music Theory Online*, vol. 23, no. 4, 2017. [Online]. Available: <https://mtosmt.org/issues/mto.17.23.4/mto.17.23.4.wells.html>
- [14] —, “Theorizing trikāla: A generalized intervallic approach to pulse transformation in south indian carnatic music,” *Analytical Approaches to World Music*, vol. 8, no. 1, pp. 48–86, 2020.

FROM RHYTHM TO REPRESENTATION: BTS'S "IDOL" AS A MULTIMODAL AUDIO SCORE

Veronika Mattová

Masaryk University

veronika.mattova@phil.muni.cz

Michal Schwarz

Masaryk University

schwarz@phil.muni.cz

ABSTRACT

Contemporary Korean music, especially through artists like group BTS, has reshaped global music by integrating traditional Korean elements with modern pop influences. This study examines how BTS's song "IDOL" functions as a multimodal audio score, blending traditional rhythmic structures like jangdan (cyclical rhythms) and dance forms such as bukcheong sajanori (lion mask dance) with contemporary K-pop production. By analyzing Jeongganbo, a traditional Korean time-based notation system, this paper demonstrates how these cultural elements are adapted and transformed into visual, rhythmic, and sonic components of "IDOL". The study highlights how rhythmic cycles, onomatopoeic syllables, and choreography are used to create a hybridized musical experience. Additionally, the research explores how "IDOL" exemplifies transmedia storytelling, where music, video, choreography, and fan interaction form a dynamic, layered narrative. This paper illustrates how BTS reinterprets traditional Korean musical practices within a global pop context, showing how multimodal audio scores and transmedia strategies allow for a deeper, more immersive engagement with music, bridging cultural and temporal gaps.

1. INTRODUCTION

K-pop, a genre of popular music coming from South Korea, a global cultural phenomenon, has revolutionized the way music is produced, consumed, and experienced worldwide. While often associated with contemporary pop music, K-pop also integrates elements of traditional Korean music (gugak, 국악), offering a hybridized musical experience (Kim et al., 2024). This research explores how BTS, one of the most successful K-pop groups, incorporates traditional Korean elements into their

work, specifically in the song "IDOL". The song and its accompanying performance serve as an excellent case study for understanding how traditional practices, such as jangdan (장단) (rhythmic structures) and bukcheong sajanori (북청사자놀이) (lion mask dance), are incorporated into modern K-pop performances, creating a unique multimodal audio score that bridges cultural and temporal gaps.

This study will analyze how multimodal representations, including audio, visual, and performative elements, function as an audio score in K-pop, particularly in Korean music band, BTS's music (Cake Corporation, 2024). By examining Jeongganbo (정간보), a traditional Korean time-based notation system, the study aims to show how elements of traditional notation are preserved, adapted, and transformed into visual and rhythmic representations in BTS's performance, ultimately creating a hybrid audio score. By situating this case study within broader frameworks of media convergence and transmedia storytelling, the paper also considers how musical meaning is distributed across artefacts, platforms, and participatory fan practices.

2. JEONGGANBO: TRADITIONAL KOREAN NOTATION AND ITS RELEVANCE TO MODERN MUSIC

Jeongganbo (정간보) is a traditional Korean musical notation system used primarily in gugak (국악) (Korean classical music) (Kim et al., 2024). Unlike Western music notation, which is pitch-based and uses a staff, Jeongganbo is organized around time segmentation and rhythmic representation (Tran, Park, & Jung, 2023). This system focuses on time-based grids rather than the representation of pitch, which makes it unique compared to Western systems of musical notation (Kim et al., 2024). Jangdan (장단), understood as a cyclical rhythmic framework, is central to Korean musical notation and performance. In their analysis of Jeongganbo, Tran, Park, and Jung (2023) demonstrate how rhythmic structure and time

Copyright: © 2024 First author et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

segmentation are encoded through repetitive, cycle-based patterns.

While Jeongganbo is not widely used today in mainstream music, its influence can still be felt in contemporary Korean musical practices, particularly in traditional performance and certain K-pop choreographies (Tran, Park, & Jung, 2023). When we look into the past again from the bigger perspective, not only Jeongganbo but also its music theory had in the Korean cultural context quite valid representation of social and cosmological order rather than a neutral technical record (Han, 1983). This higher meaning gives us another academic explanation, why is it still relevant, to this day.

3. THE ROLE OF MULTIMODAL AUDIO SCORES IN K-POP

In K-pop, music and visuals are inseparable. Music videos, lyrics, and live performances function not just as musical scores, but as multimodal representations of the music. The term multimodal audio score in context of this paper, refers to the idea that musical compositions can be represented through a combination of sound, text, image, and movement, while in the case of BTS's "IDOL", the music video (BTS, 2018) and its lyrics (Genius English Translations, 2018) serve as multimodal audio scores, where each medium adds layers of meaning to the music, allowing for a deeper, more nuanced understanding of the song.

For example, the BTS Lyrics Inside book, in its Global Edition, transforms the traditional lyrics into a hybrid score. Instead of simply presenting the text, the book incorporates rhythmic notation, onomatopoeic syllables, and diagrammatic representations of time, reflecting the structure of Jeongganbo (정간보) (Cake Corporation, 2024). The use of symbols and images to represent rhythm is an important feature of this multimodal score, as it allows the reader to engage with the music in a way that goes beyond the traditional means of music notation. The cross-modal translation between Korean onomatopoeia-creation of words that phonetically imitate or evoke sounds linked to specific actions or objects (Nissa et al., 2021) and Romanized English gloss (bum badum...) provides a rich, dynamic representation of the music, offering fans a unique way to interact with the song.

4. TRANSMEDIAL STORYTELLING ACROSS MEDIA

4.1. Media Convergence and Participatory Culture

The concept of transmedia storytelling is inseparable from Henry Jenkins' foundational work on media convergence. In *Convergence Culture: Where Old and New Media Collide*, Jenkins (2006) defines convergence not as a purely technological process, but as a profound cultural shift that reshapes how media content is produced, circulated, and consumed. Convergence refers to the flow of content across multiple media platforms, the

cooperation among different media industries, and the migratory behavior of audiences who actively seek meaningful narrative experiences across media boundaries.

Crucially, Jenkins rejects the notion of passive media consumption. Instead, he emphasizes participatory culture, where audiences function as active interpreters, collaborators, and, in some cases, co-creators of meaning. Media convergence, therefore, occurs "within the brains of individual consumers and through their social interactions with others" (Jenkins, 2006), highlighting the cognitive and social dimensions of contemporary media engagement. This participatory framework enables fans to construct personal and collective mythologies by connecting fragmented narrative elements dispersed across media environments.

Jenkins further introduces the concept of collective intelligence, adapted from Pierre Lévy, to describe how knowledge production becomes a shared social process and if we look at this topic from the convergence culture, no single individual possesses complete knowledge; rather, meaning emerges through collaboration, discussion, and shared interpretation (Lévy, 1997; Jenkins, 2006). This principle is particularly relevant to global fandoms, where interpretive labor is distributed across transnational communities and digital platforms.

4.2. Transmedia Storytelling Beyond a Buzzword

While Jenkins' work laid the theoretical groundwork for understanding convergence and participation, subsequent scholarship has critically examined how transmedia storytelling is operationalized in practice. Sousa, Martins, and Zagalo (2016) argue that transmedia storytelling has often been reduced to a fashionable buzzword, frequently misapplied to describe mere cross-media promotion or audience interaction. They contend that genuine transmedia storytelling requires the meaningful distribution of narrative elements across multiple media and artefacts, where each component contributes uniquely to the overall storyworld.

Importantly, Sousa et al. (2016) conceptualize transmedia storytelling as a dynamic process involving three principal stakeholder groups: the media industry, creators, and the public. Each group enters the transmedia ecosystem with distinct stakes. Media industries prioritize economic sustainability and brand expansion; creators seek artistic expression and narrative coherence; audiences pursue emotional engagement, identity work, and participatory belonging. These interests may diverge, but they can also converge productively when transmedia principles are authentically applied.

Rather than framing transmedia as a linear top-down strategy, Sousa et al. emphasize negotiation, reciprocity, and creative tension among participants. In this sense, transmedia storytelling becomes an ecosystem where creativity, participation, and fandom coexist alongside economic and institutional interests. This perspective extends Jenkins' participatory model by explicitly addressing power relations and the structural conditions shaping transmedia production.

4.3. K-pop as a Transmedia Ecosystem

K-pop represents a particularly illustrative case of transmedia storytelling in action. As a global cultural industry, K-pop operates through a highly integrated media environment that includes music, performance, visual aesthetics, social media, merchandise, documentaries, books, and interactive fan platforms. These elements do not merely replicate content across channels; they expand narrative meaning and deepen audience immersion. This is not some new practice. Imaginaries of all kinds can be seen in various Korean music contexts. For example, specifically instrument imagery grounds rhythmic notation in embodied Korean ensemble practice (Rockwell, 1983). This can be understood as an early form of storytelling that continues to inform imaginary practices today.

In the case of BTS, transmedia storytelling functions as a bridge between traditional Korean cultural elements and contemporary global media practices. Musical compositions incorporate traditional rhythms and symbolic references, while choreography, visual storytelling, and lyrical narratives extend these meanings across formats (Cook, 2001). This convergence between “old” and “new” media reflects Jenkins’ assertion that convergence culture does not eliminate older media forms but reconfigures their relationships within new contexts.

From the perspective proposed by Sousa et al. (2016), BTS’ transmedia universe exemplifies a relatively balanced alignment of stakeholder stakes. The media industry (HYBE- Company of BTS) strategically coordinates platforms and distribution; the artists actively participate in narrative construction and thematic continuity; and the fandom (ARMY) engages in interpretive, creative, and affective labor that sustains the storyworld across cultural and linguistic boundaries.

4.4. The BTS Lyrics Inside Book as a Transmedial Artefact

The BTS Lyrics Inside book can be understood as a transmedial artefact that crystallizes the convergence of music, text, and visual design. Rather than functioning solely as a paratextual supplement, the book recontextualizes lyrics as literary and cultural material, inviting close reading, translation, and reinterpretation. This artefact enables fans to engage with the music beyond auditory consumption, transforming listening into a multimodal interpretive practice.

Through this lens, fans become active participants in meaning-making processes, aligning with Jenkins’ model of participatory culture and Sousa et al.’s emphasis on fandom as a central transmedia actor. The Lyrics Inside book operates as a node within a larger narrative network, connecting albums, performances, interviews, and fan-generated content. It thus supports a shared narrative space where emotional symbolism, cultural identity, and collective intelligence intersect, far behind the music, itself.

5. DISCUSSION: CASE STUDY

5.1. BTS “IDOL” as a Multimodal Audio Score

The BTS Lyrics Inside book offers a compelling example of how elements of traditional Korean music are recontextualized through contemporary multimodal notation. Rather than functioning solely as a repository of lyrics, the book operates as a hybrid audio score that combines text, sound cues, visual layout, and performance imagery into a single interpretive space.

Through vocalized rhythmic patterns such as “덩기덕 쿵더러러” (deonggi-deok kung-deoreo, onomatopoeic syllables imitating drum sounds) (Cake Corporation, 2024), cross-modal translations, for example from Korean onomatopoeia to Romanized English glosses, and diagrammatic layouts, the act of reading lyrics is transformed into a multimodal experience (Boddy & Boddy, 2010). Images of performance further extend this experience by visually referencing bodily movement and rhythmic gesture, suggesting how sound, motion, and meaning coalesce in performance (Cake Corporation, 2024).

Importantly, this multimodal strategy does not merely illustrate the song but actively guides interpretation. The reader is positioned not as a passive consumer of lyrics but as an embodied participant who is invited to imagine rhythm, movement, and affect simultaneously.

The lyrics explicitly include traditional exclamations, for instance:

“열쌍 좋다 ... 지화자 좋다”

(yeolssang jota ... jiwaja jota,

“Oh yeah, it’s awesome ... Hooray, it’s so awesome”)

(Cake Corporation, 2024)

As explained in the BTS Lyrics Inside book, 열쌍 (yeolssang, a joyful exclamation) and 지화자 (jiwaja, a celebratory chant) are vocal expressions known as 추임새 (chuiimsae, supportive rhythmic exclamations), traditionally used during singing or dancing in Korean music. In “IDOL,” the phrase “You can’t stop me lovin’ myself” followed by these exclamations functions as an expression of heightened excitement embedded directly into the rhythmic structure of the song (Cake Corporation, 2024).

Beyond their explanatory role, these exclamations also signal a continuity between traditional performance practices and contemporary pop expression. Their inclusion suggests that affective participation, collective response, and vocal embodiment remain central to musical meaning-making across genres and historical periods (Catalano, 2024).

This multimodal audio score enables fans to engage with BTS’s music beyond listening alone, encouraging exploration of traditional jangdan as rhythmic cycle

principles that underlie the song. The integration of these rhythmic frameworks into a global pop format illustrates the hybrid nature of K-pop, where inherited musical logics coexist with contemporary production and global circulation.

This hybridity becomes especially visible in live performance contexts. In the 2018 Melon Music Awards (MMA) special performance of “IDOL,” the members of BTS incorporated traditional Korean dance forms as part of a large-scale pop spectacle. As documented in the lyrics book, a drum dance, a fan dance, and a mask dance, drawing on performance traditions such as pungmulnori (풍물) traditional percussion-based folk performance) and bukcheong sajanoreum (북청사자놀이), also known as alion mask dance ritual (Cake Corporation, 2024), used as a valid cultural component in nowadays vibrating performance.

Here, choreography functions as an extension of notation. Movement becomes another representational layer through which rhythm and cultural reference are communicated, reinforcing the idea of the song as a multimodal score that unfolds across media and performance contexts.

This logic extends into the music video itself. Beyond the lyrics, the choreography, costumes, and visual aesthetics incorporate recognizable elements of Korean cultural heritage. These include lion mask dancing from lion mask dance and also buchaechum (부채춤), in our context knowns as a fan dance- motifs, and visual references to Korean landscape painting. These elements are juxtaposed with both traditional and Euro-Asian architectural imagery, emphasizing the layered and hybrid visual language of “IDOL” (Cake Corporation, 2024).

Such visual strategies further underscore how cultural memory is activated through imagery. Rather than functioning as static references, these elements invite imaginative engagement, allowing viewers to connect rhythm, movement, and symbolism across historical and contemporary frames.

Finally, rhythmic mnemonics such as “덩기덕 쿵더러러” (deonggi-deok kung-deoreo, vocalized representation of drum strokes) (Figure 1) imitate the sounds of the 장구 (janggu, hourglass-shaped traditional Korean drum). As described in the lyrics book, these vocalizations correspond to specific drum strikes within traditional jangdan (rhythmic cycles) (Cake Corporation, 2024).

⊙		○	
명 (deong)	기덕 (gi-deok)	쿵 (kung)	더러러러 (deo-reo-reo-reo)

Figure 1. “IDOL’s” “덩기덕 쿵더러러” (deonggi-deok kung-deoreo) as commonly used jangdan in traditional Korean music. This notation represents rhythmic patterns from Jeongganbo, a traditional Korean time-based system. Adapted from Cake Corporation. (2024). BTS Lyrics Inside (Global Edition). Big Hit Music Co., Ltd. ISBN 979-11-90996-80-8.

In this sense, “IDOL” (BTS, 2018) exemplifies how traditional Korean music notation and performance logic can be reimagined as a multimodal, performative score. Here, one can already observe early narrative and symbolic structures that continue to support diverse imaginary practices in contemporary popular culture.

6. LIMITATIONS AND FUTURE RESEARCH

This study is based on a single case study, BTS’s “IDOL,” which allows for a detailed analysis but may appear selective. While the song was chosen due to its explicit integration of traditional Korean rhythmic structures, performance practices, and visual symbolism, the findings should not be generalized to all K-pop productions.

Future research aims to expand this collaborative Korean music research to include multiple artists, groups, and songs that employ similar musical and performative strategies. Examining a broader corpus would enable comparative analysis and strengthen the generalizability of the proposed concept of multimodal audio scores.

Such expansion may also benefit from interdisciplinary approaches, combining musicology, media studies, and multimodal analysis, to further explore how traditional Korean musical frameworks continue to inform contemporary popular music.

7. CONCLUSION

The integration of traditional Korean music (gugak, 국악) into contemporary K-pop offers a compelling example of multimodal music representation, in which sound, text, image, notation, and performance interact across media. In this context, music emerges not merely as an auditory phenomenon but as a distributed representational system spanning multiple artefacts and platforms (Han, 2018).

Korean music notation, particularly Jeongganbo (정간보), provides an important conceptual background for this discussion. As a time-based notational system that prioritizes rhythm, duration, and embodied performance, Jeongganbo reflects a long-standing understanding of music as a performative and interpretive practice rather than a fixed textual object. This historical perspective highlights that notation in Korean musical contexts has traditionally functioned as a mediating framework between sound, body, and meaning.

Against this backdrop, the *BTS Lyrics Inside* book can be understood as a contemporary hybrid audio score. Through rhythmic mnemonics, visual layouts, and instrumental imagery, it renders rhythmic structure legible without relying on conventional staff notation. While these representations do not constitute formal notation systems, they operate according to analogous principles, translating rhythm, timing, and gesture into multimodal forms. In doing so, popular music artefacts implicitly echo earlier Korean notational logics while adapting them to global and transmedia environments.

From a transmedial perspective, musical meaning circulates across albums, lyrics books, music videos, merchandise, and live performances. These artefacts do not merely accompany sound but actively participate in its representation and interpretation, enabling fans to engage with music as a visual, textual, and embodied experience. This participatory dimension underscores the role of fans as interpreters within contemporary K-pop ecosystems.

These observations align with current research in Music Information Retrieval and music computing, which increasingly emphasizes the importance of multimodal representations for understanding musical content (Simonetta et al., 2019). Although this study does not propose computational models, it offers a culturally grounded case illustrating how multimodal music representation already operates in practice.

Looking ahead, hybrid artefacts such as lyrics books, annotated visual scores, and interactive media environments represent promising sites for future research at the intersection of Korean music notation, multimodal representation, and transmedia storytelling. Examining these forms may offer valuable insights into how music is not only heard but also seen, read, and embodied in the digital age, contributing to a broader understanding of sound as score.

Acknowledgments

This paper was prepared thanks to financial support within the project “Evolutionary Interferences of Religion and Governance in Inner Asia: Comparison of mutual impacts with tributary countries: Mongolia, Korea, and Vietnam” financed by the Czech Science Foundation at Masaryk University, Faculty of Arts (project code: GA23-06953S).

8. REFERENCES

1. Boddy, J., & Boddy, J. (2010). A multimodal way of experiencing and exploring music. *Interdisciplinary Science Reviews*, 35(2), 138–153.
2. BTS. (2018, August 24). *IDOL* [Music video]. YouTube. <https://www.youtube.com/watch?v=pBuZEGYXA6E>
3. Cake Corporation. (2024). *BTS lyrics inside* (Global edition). Big Hit Music Co., Ltd. ISBN 979-11-90996-80-8
4. Catalano, E. (2024). The colonial era and beyond: The past and present and its effect on music and K-pop in Korea. *Colorado Journal of Asian Studies*, 11(1). <https://doi.org/10.33011/cjas.v1i1.2963>
5. Cook, N. (2001). Theorizing musical meaning. *Music Theory Spectrum*, 23(2), 170–195. <https://doi.org/10.1525/mts.2001.23.2.170>
6. Genius English Translations. (2018, August 24). *BTS – IDOL* (English translation). Genius. <https://genius.com/Genius-english-translations-bts-idol-english-translation-lyrics>
7. Han, M.-Y. (1983). Religious origins of Korean music. In Korean National Commission for UNESCO (Ed.), *Traditional Korean music* (pp. 23–34). The Sisa-yong-sa Publishers.
8. Han, D., Gotham, M., Kim, D., Park, H., Lee, S., & Jeong, D. (2024). Six dragons fly again: Reviving 15th-century Korean court music with transformers and novel encoding [Preprint]. arXiv.
9. Han, D.-Y. (2018, March). Gugak and K-pop: Sharing the spotlight. Korean Cultural Center. <https://www.korean-culture.org/eng/webzine/201803/sub03.html>
10. Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York University Press.
11. Jung, J., Kim, D., Lee, S., Cho, S., Soh, H., Bukey, I., Donahue, C., & Jeong, D. (2025). Unified cross-modal translation of score images, symbolic music, and performance audio (arXiv:2505.12863). arXiv. <https://doi.org/10.48550/arXiv.2505.12863>
12. Kim, D., Han, D., Jeong, D., & Valero-Mas, J. J. (2024). On the automatic recognition of Jeongganbo music notation: Dataset and approach (Version 1) [Preprint]. Research Square. <https://doi.org/10.21203/rs.3.rs-4668854/v1>
13. Lévy, P. (1997). *Collective intelligence: Mankind's emerging world in cyberspace* (R. Bononno, Trans.). Perseus Books.
14. Nissa, F. K., Herniwati, Triarisanti, R., & Azizah, A. (2021). Forms and translation procedures of Korean onomatopoeia and mimesis in the webtoon. In *Proceedings of the Fifth International Conference on Language, Literature, Culture, and Education*

(ICOLLITE 2021). Atlantis Press.
<https://doi.org/10.2991/assehr.k.211119.046>

15. Rockwell, C. J. (1983). Trends and developments in Korean traditional music today. In Korean National Commission for UNESCO (Ed.), *Traditional Korean music* (pp. 106–117). The Si-sa-yong-sa Publishers.
16. Simonetta, F., Ntalampiras, S., & Avanzini, F. (2019). Multimodal music information processing and retrieval: Survey and future challenges. In *Proceedings of the 2019 International Workshop on Multilayer Music Representation and Processing (MMRP)*.
<https://doi.org/10.1109/MMRP.2019.8665366>
17. Sousa, M. N. e, Martins, M. de L., & Zagalo, N. (2016). Transmedia storytelling: The roles and stakes of the different participants in the process of a convergent story, in divergent media and artefacts. In A. Lugmayr & C. Dal Zotto (Eds.), *Media convergence handbook* (Vol. 2, pp. 117–135). Springer. https://doi.org/10.1007/978-3-642-54487-3_7
18. Tang, M., Li, J., Yang, L., Zhang, Z., Tian, J., Li, Z., Zhang, L., & Wang, P. (2025). NOTA: Multimodal music notation understanding for visual large language model (arXiv:2502.14893). arXiv.
<https://doi.org/10.48550/arXiv.2505.12863>
19. Tran, M. L., Park, C., & Jung, J. H. (2023). Topological data analysis of Korean music in Jeongganbo: A cycle structure. *Journal of Mathematics and Music*, 17(3), 403–432.
<https://doi.org/10.1080/17459737.2022.2164626>

INDIVIDUALISING GRAPHIC NOTATION THROUGH BOTANICAL METAPHORS: A PRACTICE-BASED STUDY OF *BOTÁNICA* (2024)

Jingsong Teng

Independent composer; Royal Danish Academy of Music (alumnus), Copenhagen
jingsong6667@gmail.com

ABSTRACT

Botánica (2024) is a three-part suite for five to eighteen pitched instruments in graphic notation. Across its movements—*Tree*, *Flower*, and *Ivy*—the score translates botanical growth processes into performable rule sets, so that “nature” functions not only as imagery but as an interface for ensemble decision-making. In this sense, the score operates as a listening guide: it specifies where decisions occur and how sonic states are maintained, rather than prescribing fixed musical events. This study builds on practice-as-research and draws on Peircean semiotics as one lens for understanding how the score is read. I introduce the concept of individualised graphic notation, in which each movement uses a distinct visual grammar to distribute performer choice across parameters (timing, path selection, local ordering) while retaining a clear sonic identity. Through analysis of the score and rehearsal documentation, I show that the piece explores three degrees of aleatoric control in practice. *Tree* employs branching layers that allow flexible pacing within a fixed pitch world. *Flower* uses adjacency constraints (inspired by phyllotaxis) to yield emergent harmonies. *Ivy* combines limited transposition with shape-coded ordering freedoms to model gradual transformation. I argue that ecological metaphors become compositionally useful when formalised as constraints that performers can read, negotiate, and collectively maintain.

1. INTRODUCTION

Graphic scores often invite performers to co-author the musical work, yet the practical question remains: how can a composer write “openness” without losing identity, coherence, and ensemble coordination? *Botánica* emerged from this question and from a personal fascination with plant life cycles—germination, proliferation, and decline—as a model of structured variability. Rather than attempting to represent plants sonically, I aimed to translate aspects of botanical behaviour into rules for musical motion. In this sense, the work belongs to traditions of open works and

experimental notation (e.g., Cage, Brown, Cardew, Ferneyhough), but it also treats the visual surface of the score as an ecological diagram: a map of possible actions rather than a fixed script. This approach parallels Ferneyhough’s Lemma–Icon–Epigram model [1] and aligns with recent scholarship that views graphic scores as image texts [2] or as managed-choice systems [3].

This article presents an individualised graphic notation approach in which each movement employs a distinct visual grammar to invite different kinds of performer decisions. It also proposes a simple framework (summarised in Table 1) for comparing degrees of openness by identifying what in the music is free, fixed, or negotiated in real time, and it documents how rehearsal feedback led to notation and performance-note revisions that clarify performers’ choices without narrowing interpretive space.

2. FRAMEWORK AND METHOD

I work within a practice-as-research framework [4]: composition functions as a method of inquiry, and the artefact (score + performance) is treated as evidence. To describe how the score communicates, I draw on Peircean [5] semiotics, where signs can function iconically (resemblance), indexically (causal trace), or symbolically (convention). Graphic notation typically mixes all three: drawings may resemble motion, arrows may index direction, and staff fragments invoke conventional pitch reading. Thinking semiotically helps clarify which parts of the score aim for intuitive “reading” and which require explicit instruction.

Materials considered here include: (a) the final scores; (b) documented rehearsal observations and performer questions; (c) the post-rehearsal adjustments I made to the notation and performance notes. Where relevant, I describe revisions as clarifications (making an intended freedom legible) rather than constraints (removing freedom).

2.1. Practice-based evidence and documentation

In addition to the score itself, I treat moments of misunderstanding as useful data: when performers ask whether a choice is permitted, or when an ensemble converges on a single reading, those reactions reveal which freedoms are actually legible. I therefore document (i) recurrent performer questions, (ii) points where

Copyright: © 2026 Jingsong Teng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Table 1. Framework of openness used in rehearsal, indicating fixed, variable, and negotiated elements.

Movement	Variable elements	Fixed elements	Negotiated elements	Rehearsal focal points
<i>Tree</i>	tempo/pacing; articulation; micro-timing; path selection	pitch world; branching layout; overall time span	density/balance; transition timing; shared time reference	Transition timing, desynchronised movement, and density balance.
<i>Flower</i>	adjacent-petal path choice; local pitch selection	adjacency rule; pitch distribution; overall time span	harmonic continuity; registral continuity	Adjacency rules, harmonic continuity, and ending coordination.
<i>Ivy</i>	ordering within shapes; reversals/permutations	pattern duration; transposition plan; shared arrival points; overall time span	controlled divergence; synchronised arrivals	Shape permissions, variable parameters, and shared arrivals.

ensemble coordination breaks down, and (iii) the specific notational edits made in response. This produces a traceable chain from design intention → rehearsal behaviour → revision.

2.2. Semiotic design of the score surface

Peirce’s triad (icon-index-symbol) is useful here because *Botánica* deliberately mixes reading modes. Branching diagrams and petal layouts operate iconically (they resemble growth forms), arrows and gradients work indexically (they point to direction/energy), and pitch labels, legends, and staff fragments function symbolically (they rely on learned conventions). A practical aim of my revisions is to keep the semiotic mix stable inside each movement: when an element is intended to be ‘read by intuition’ (icon/index), I avoid over-explaining it; when an element requires convention (symbol), I make the rule explicit at the point of use [5]. This also aligns with semiotic accounts of open notation that treat ambiguity as a compositional resource [6].

2.3. Parameter map of openness

To make the degrees of openness comparable, I describe each movement by which parameters are (a) free, (b) fixed, and (c) negotiated in real time. Table 1 summarises this map. The goal is not to quantify freedom, but to make the locus of agency explicit so that the design can be transferred to other works (see Table 1).

Taken together, these observations lead me to treat the score not simply as a visual object, but as a practical environment for decision-making. In *Botánica*, graphic notation does three kinds of work at once: it encodes musical information, it distributes agency among performers, and it shapes how those choices become legible in rehearsal and performance. This is why I describe the piece as an instance of individualised graphic notation. Each movement uses a different visual grammar not for visual variety alone, but to organise a different relation between what is free, what is fixed, and what must be negotiated in real time. From this perspective, openness is not a re-

duction of control. It depends on placing constraints, options, and cues with enough precision that performers can act independently without losing the movement’s sonic identity. Rehearsal is therefore part of the compositional method: moments of uncertainty show where intended freedoms are not yet legible, and subsequent revisions help the score function more clearly as both notation and listening guide.

3. WORK OVERVIEW

Botánica is written for five to eighteen pitched instruments (any instrumentation, mixed timbres). The flexible instrumentation supports the work’s ecological premise: the same score can be “inhabited” by different ensembles, producing different balances and micro-histories. Although the suite is open to 5–18 players overall, *Flower* tends to register most clearly with a smaller group. The number range also alludes to the 18-month structure of the Mayan Haab’ calendar and its additional five Wayeb’ days; within the piece, these numbers function as a compositional frame for cyclical time rather than as a direct act of cultural representation. [7]

Across all three movements, performers keep a shared sense of duration using a timer, and one player gives an initial cue to begin together. Synchronised endings are not required; instead, the ensemble “breathes” around overlapping gestures, encouraging a texture where individual trajectories remain audible.

3.1. Performance protocol

Because the ensemble may vary in size and instrumentation, *Botánica* relies on a small set of shared protocols. In rehearsal I introduce them explicitly to reduce ‘hidden assumptions’ about how open scores should behave:

- Use a timer or a stopwatch to keep a common sense of large-scale duration; local timing remains flexible.

- Appoint a single cue-giver only for the start (and, if needed, a mid-point cue in *Flower*); the rest is coordinated by listening.
- Treat balance as relational rather than absolute: adjust register, timbre, and density so that no voice dominates unless intentionally foregrounded.
- Read the legend at the moment of use: when a symbol encodes a rule (symbolic sign), it must be encountered where the decision occurs.

4. MOVEMENT CASE STUDIES

4.1. *Tree*: branching layers and controlled pacing

Tree is realised through a navigation protocol in which players may diverge at branch points while maintaining a shared temporal frame, negotiating density and timing by listening.

- Choose one seed material and begin in a shared dynamic ‘low-energy’ state.

- At each branch point, select the next pattern; players may choose different routes.
- Keep pitch identity stable, but allow pacing to stretch or compress — especially when density becomes high.
- Use the right-edge dynamic arc as the macro-shape, not as bar-by-bar control.

Tree presents a branching diagram built from several layers of notational information. At the base, performers choose one of four “seed materials.” From there, they traverse upward through branch points, selecting subsequent patterns while following a broad dynamic arc indicated at the right edge. The principal freedom lies in pacing and articulation: players may linger, compress, or expand local durations, but the pitch world and the general energetic trajectory remain stable. The result is a layered soundscape that can fluctuate in density without losing direction.

1. TREE - for 5 - 18 any pitched instruments

(The score is in C)
Senza misura

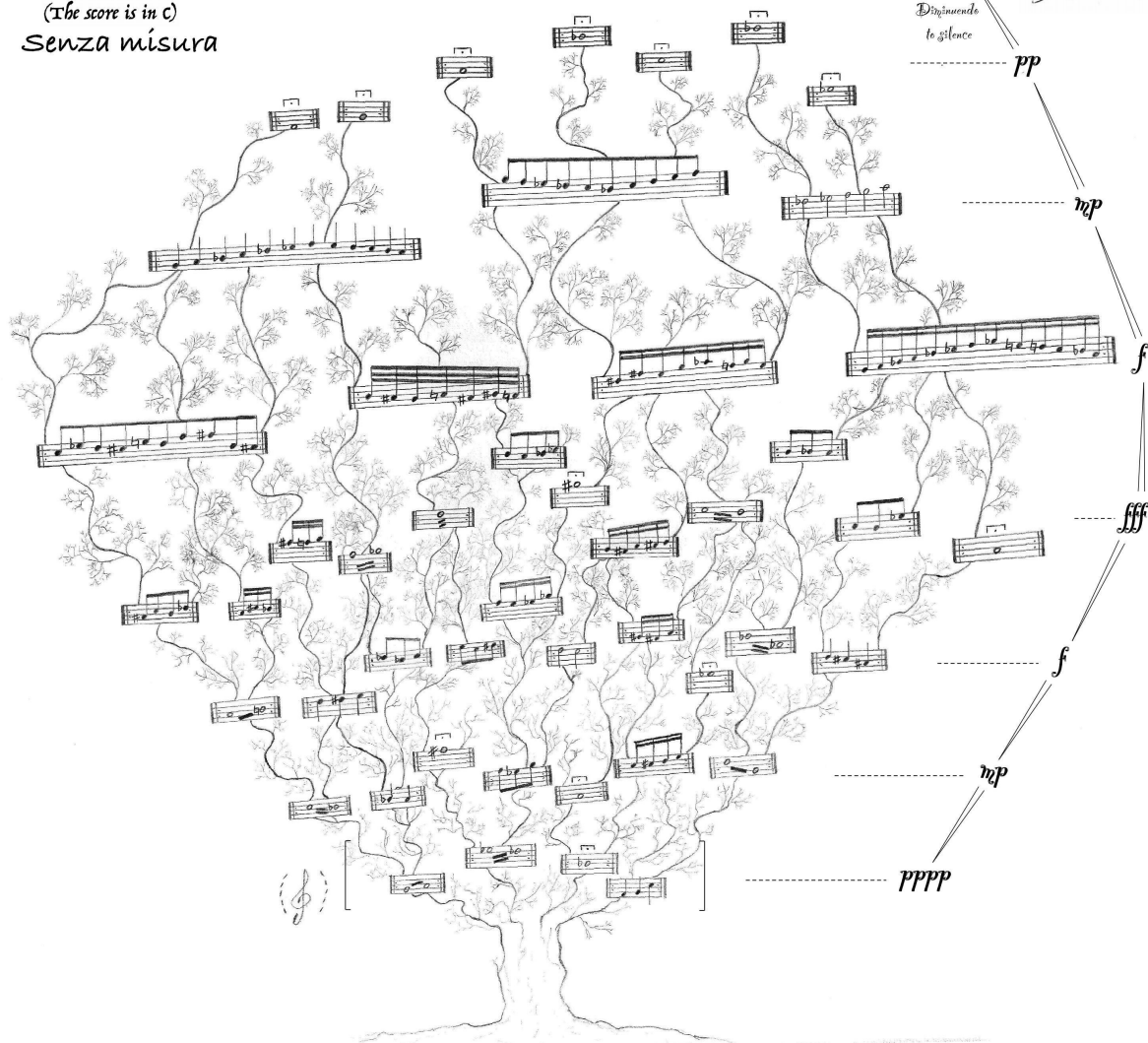


Figure 1. *Tree*: score showing branching layout and layer cues.

The pitch material in *Tree* is intentionally limited. By keeping pitch relations familiar and relatively consonant, the notation can afford greater rhythmic and timbral variability without collapsing into noise. In rehearsal, this limitation proved useful: performers could focus on navigation and ensemble listening rather than on decoding complex pitch sets. At the same time, the branching layout encourages micro-differences between players, so that the “same” ascent can sound like many concurrent growth lines.

4.2. *Flower*: adjacency rules and emergent harmony

In performance, this constraint is realised as a simple navigation protocol: performers begin at the centre and move only to adjacent petals, preserving local harmonic neighbourhoods.

- Start in the centre; each subsequent move must go to a neighbouring petal, with ‘neighbour’ defined clearly in the legend.
- Think locally: avoid long ‘teleport’ jumps that would erase the intended harmonic neighbourhood.
- Let harmony emerge from shared proximity: recurring sonorities are a feature.
- Follow the performance indications for changing dynamics and petal durations, so the movement reads as a single process rather than a collection of chords.

Flower shifts from linear growth to radial expansion. The score is arranged as petals around a centre. Players begin in the centre and then move only to petals adjacent to their current location. This simple adjacency constraint draws inspiration from phyllotaxis, where the placement of new organs is shaped by what is already nearby. Musically, the rule prevents performers from jumping freely across the field; instead, harmony emerges from locally constrained wandering. [8,9]

Pitch classes are distributed unevenly across the petals. Some pitches recur more frequently, creating a subtle gravitational pull (a quasi-tonal centre in G minor) even though no fixed harmonic progression is prescribed (see Figure 2). Because performers are restricted to adjacent-petal choices, the harmonic field develops through locally connected pitch groupings rather than large registral leaps or freely scattered note selection. I therefore realise *Flower* with a smaller ensemble (around five players), so that recurring sonorities and the subtle pull toward G minor can register more clearly.

The movement compresses a life-cycle arc into changing dynamics and petal durations. It begins near stillness (*ppp-p*), intensifies through shorter petal durations and higher dynamics (*mf-f*), and finally collapses toward quiet, suggesting decline. Here the “botanical” is not only an image: it is encoded as a set of time-dependent constraints that performers can read and collectively shape.

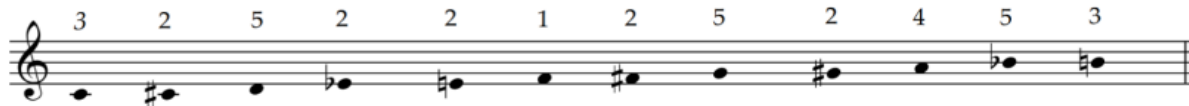


Figure 2. Distribution of pitch classes in *Flower*.



Figure 3. Two examples of possible playing orders in *Flower*.

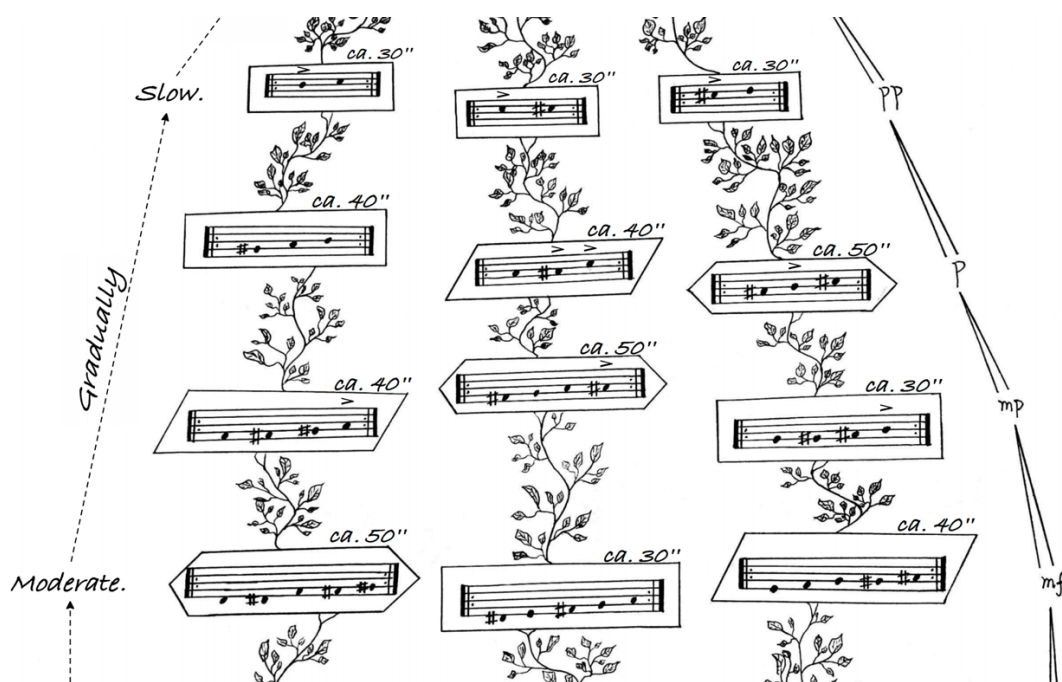


Figure 4. *Ivy*: Excerpt showing shape-coded patterns and transposition steps.

4.3. *Ivy*: limited transposition and shape-coded ordering freedoms

Ivy treats the transposition process as a stable spine: within each step, players reorder shape-coded materials locally, while longer-range coherence is maintained through the fixed transformation plan and shared arrival points.

- Treat the transposition process as the ‘stem’ that persists throughout the movement.
- Apply shape rules consistently: when a shape grants reversal or permutation, use it to create small divergences between players.
- Allow temporal divergence inside a layer, but aim to recognise shared endpoints as convergence points.
- If coordination becomes unclear, prioritise the macro-direction (drift through transposition) over synchronising details.

Ivy focuses on gradual transformation through repeated material under systematic change. The pitch framework is derived from Messiaen’s modes of limited transposition, treated as a reservoir for several short sequences. Across the piece, these sequences are transposed by semitone steps; at the same time, higher layers contain fewer sequences, so texture thins as transposition accumulates. This pairing of repetition and drift models *ivy*’s climbing habit: persistence with adaptation. [10,11]

In *Ivy*, graphic shapes indicate different degrees of freedom in ordering: rectangles require strict left-to-right reading; parallelograms allow reversal; hexagons permit any internal order. Within a single layer, multiple paths have unequal durations yet converge at a shared endpoint, so performers experience temporal divergence inside a

collective frame. In rehearsal, this asymmetry reliably produced a feeling of “searching”: lines move at different speeds while remaining oriented toward the same arrival point.

5. PERFORMANCE-INFORMED REVISIONS

For open notation to function as intended, performers need to understand clearly what they are allowed to do. As a result, several of my revisions focused not on changing the musical content, but on clarifying which choices were open and which constraints needed to remain stable. Across rehearsals, three recurring issues emerged:

- Early commitment to a single route. Some players settled on one path too quickly and tended to repeat it, which reduced the intended variability. I therefore added explicit encouragement to vary paths and to treat repetition as a choice rather than a default.
- Dynamic balance in variable ensembles. With flexible instrumentation, balance cannot be specified in absolute terms. I therefore rewrote the dynamic instructions in relational terms (listen, adjust, keep every voice audible), which performers found clearer.
- Shape meanings not immediately clear from the legend. In *Ivy*, some shape-based freedoms were not immediately legible in rehearsal. I therefore clarified the legend and its wording so that the permitted choices were more explicit.

In fact, this suggests that providing clarity doesn’t diminish openness; instead, being explicit about options can enhance variety by preventing performers from unwittingly limiting themselves.

Table 2. Revision log: clarifications that support performer agency.

Observed issue	Revision made	Effect on openness
Early commitment to a single route	Add explicit permission to vary or avoid fixed paths	Prevents unintended self-restriction; increases variety without adding material
Dynamics treated as absolute levels	Rewrite dynamics as relational instructions	Adapts across instruments and rooms; supports ensemble agency
Shape meanings not immediately clear from the legend	Clarify the legend and its wording	Reduces uncertainty around rule-based symbols; preserves freedom through clarity

6. DISCUSSION: INDIVIDUALISING GRAPHIC NOTATION

In *Botánica*, openness is organised differently in each movement by assigning performer choice to specific parameters. *Tree* is open primarily in pacing and articulation; *Flower* is open in path choice under adjacency constraints; *Ivy* is open in local ordering within a tightly controlled transformational process. Described this way, the three movements form a small comparative set: not three stylistic variations, but three configurations of agency.

The botanical metaphor matters because it is formalised. When “growth” is translated into constraints that performers can enact — branching decisions, neighbour rules, convergence endpoints—the metaphor becomes an operator. This helps avoid a common pitfall in eco-themed composition: nature as surface narrative without practical consequence. Here, ecological thinking is embedded in how the ensemble negotiates time, proximity, and coordination.

However, there are limitations to this approach. The work’s flexibility also creates variability in rehearsal needs: some ensembles may require more time to internalise the navigation rules. Projection quality and page layout affect readability; future iterations may benefit from a digital score interface that can highlight the relevant rule at each moment.

Although *Botánica* is not a software project, its parameter map and rehearsal-driven revisions suggest one possible way of thinking about representation tools. In a digital score interface, the “free / fixed / negotiated” categories could be reflected as explicit constraints and optional choices, while coordination prompts could appear as brief, local annotations at decision points.

Rather than proposing a general platform, the paper suggests a practical way of translating compositional openness into interface elements that support ensemble navigation.

7. CONCLUSION

By treating botanical processes as performable constraints, *Botánica* develops an approach to graphic notation where each movement’s visual design is inseparable from its musical agency. The concept of

individualised graphic notation offers one way to compose openness with identity: not by adding more information, but by tailoring the score’s semiotic mix—icon, index, symbol—to the kind of decision the performer must make. Future work may extend this approach toward digital and real-time score systems, where ecological data and performer interaction could reshape the notation during performance. This also encourages an ecological mode of listening that treats music as interaction with a sounding environment [12,13].

Beyond this case study, the framework may be useful in pedagogy and rehearsal practice as a listening guide: it makes agency explicit, helps ensembles agree on coordination rules, and turns moments of misunderstanding into actionable score edits. Future work could test the framework in other ensembles and notational media (including audio-score layers), and examine how different cueing strategies shape collective listening and perceived form.

SUPPLEMENTARY MATERIAL

A performance video of *Botánica* (2024) is [available here](#).

REFERENCES

- [1] B. Ferneyhough, *Lemma-Icon-Epigram*. London: Edition Peters, 1981.
- [2] J. Metcalf, “What Do Graphic Scores Want? Imagetext, Indeterminacy, and Iconoclasm,” *Music in Art*, vol. 48, nos. 1-2, pp. 175-186, 2023.
- [3] L. Brøndum, “Graphic Notation, Indeterminacy and Improvisation: Implementing Choice Within a Compositional Framework,” *Open Cultural Studies*, vol. 2, pp. 639-653, 2018, doi: 10.1515/culture-2018-0058.
- [4] G. Sullivan, *Art Practice as Research: Inquiry in Visual Arts*. Thousand Oaks, CA: SAGE, 2010.
- [5] C. S. Peirce, *Collected Papers of Charles Sanders Peirce*, vol. 2: *Elements of Logic*. Cambridge, MA: Harvard University Press, 1931.

- [6] T. McKay, *A Semiotic Approach to Open Notations: Ambiguity as Opportunity*. Cambridge: Cambridge University Press, 2021.
- [7] H. Kettunen and C. Helmke, *Introduction to Maya Hieroglyphs* (17th ed.). Bonn: Wegweiser zur Maya-Schrift, 2020.
- [8] C. Kuhlemeier, "Phyllotaxis," *Current Biology*, vol. 27, no. 17, pp. R882-R887, 2017, doi: 10.1016/j.cub.2017.05.069.
- [9] D. Reinhardt, "Regulation of Phyllotaxis," *The International Journal of Developmental Biology*, vol. 49, pp. 539-546, 2005.
- [10] E. Gianoli, "The behavioural ecology of climbing plants," *AoB PLANTS*, vol. 7, plv013, 2015, doi: 10.1093/aobpla/plv013.
- [11] E. Gianoli, "Evolution of a Climbing Habit Promotes Diversification in Flowering Plants," *Proc. R. Soc. B*, vol. 271, no. 1552, pp. 2011-2015, 2004, doi: 10.1098/rspb.2004.2827.
- [12] M. Reybrouck, "A Biosemiotic and Ecological Approach to Music Cognition: Event Perception Between Auditory Listening and Cognitive Economy," *Axiomathes*, vol. 15, no. 2, pp. 229-266, 2005, doi: 10.1007/s10516-004-6679-4.
- [13] L. Lim, "Composition and Ecological Listening," in *The Cambridge Companion to Composition*, T. Young, Ed. Cambridge: Cambridge University Press, 2024.

SIGN/E: SYMBOL-BASED INTERFACE FOR GRAPHIC NOTATION EDITION FOR ABLETON LIVE

Pierre-Luc Lecours

Université de Montréal
Laboratoire formes • ondes
(LFO), CIRMMT
pierre-luc.le-
cours@umontreal.ca

Nicolas Bernier

Université de Montréal
Laboratoire formes • ondes
(LFO), CIRMMT, Hexagram
nicolas.bernier@umontreal.ca

Evan Montpellier

Université de Montréal
Laboratoire formes • ondes
(LFO)
evan.montpellier@gmail.com

Philippe-Aubert
Gauthier

Université du Québec à
Montréal
CIRMMT, Hexagram
gauthier.philippe-aubert@uqam.ca

ABSTRACT

This article outlines the functionality of SIGN/e (Symbol-based Interface for Graphic Notation Edition), a set of Max for Live devices allowing to easily create, read and animate graphic scores within Ableton Live. It addresses the need for an efficient DAW-integrated notation system suited for electronic, instrumental and mixed-media practices. Taking advantage of the intuitive Ableton Live workflow, SIGN/e allows the creation and parametric control of individual symbols and textual elements as well as management of global visualisation properties. This architecture allows the construction of dynamic synchronized graphic scores while benefiting from Ableton Live's timeline and automation system. SIGN/e is built entirely with standard Max objects to ensure long-term stability and compatibility. The visual rendering pipeline relies on Jitter and OpenGL, with shaders written in Gen to handle specific needs such as grid generation, cropping, and compositing. SIGN/e provides a flexible environment for writing and performing animated scores. This article summarizes SIGN/e's conceptual foundations, technical structure, and potential for expanding contemporary notation practices.

1. INTRODUCTION

Technological and computational advances over the past decades have led to the development of numerous software tools dedicated to the creation, editing, and playback of musical scores, most of them revolving around Common Western Notation (CWN), while offering limited options for idiosyncratic graphic score creation.

Copyright: © 2025 Pierre-Luc Lecours, Nicolas Bernier, Evan Montpellier & Philippe-Aubert Gauthier. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹ Please also see this video summarizing SIGN/e functionality: <https://youtu.be/WrAxyXMXBoQ>

SIGN/e (Symbol-based Interface for Graphic Notation Edition) is designed for the creation, animation, and playback of graphic scores within Ableton Live. SIGN/e consists of a set of AMXD devices that allow users to import, modify, move, and animate graphic symbols (raster image files like JPGs, TIFFs and PNGs) in a 2D space that scrolls horizontally in sync with Ableton Live's timeline. The device is the outcome of several iterations of research-creation projects focused on notation for Ensemble d'oscillateurs [1]. Since 2016, the ensemble's activities have contributed to electronic music group performance practices, the development of graphic notation systems, and the design of digital tools dedicated to the transmission of electronic music [2]. Notably, the ensemble previously led to the development of Graphic Score Reader (GSR), a device enabling the playback of scrolling graphic files in Ableton Live. The conception and development of SIGN/e build on this knowledge. This article will first outline a brief overview of contemporary software initiatives in graphic, animated, and interactive notation to contextualize SIGN/e's design history, technical specificities and functionality¹.

2. SCORE NOTATION EDITOR AND READER OVERVIEW

There are several notation software options available; the most well-known, such as Finale, Sibelius, MuseScore, and Dorico, are specializing in CWN, offering only limited options for graphic score creation.

Other initiatives, like OpenMusic [3] and LilyPond [4], offer more flexible and programmable approaches to composition and musical notation. OpenMusic employs object-oriented programming, while LilyPond relies on a textual syntax to produce scores that meet professional quality standards. While these applications offer options to output graphical elements, their focus remains rendering CWN scores.

A small number of plugins that allow graphical notation symbols have been integrated into music production environments, such as MaxScore2 [5], which operates within Ableton Live and Max. MaxScore2 supports graphical notation, interactive scores and CWN through the BACH library [6], another Max package. BACH provides advanced computer-assisted composition features, including graphic notation and real-time score generation.

Other score-reading applications have emerged to encompass the diversity of notation practices, including non-conventional notation. The simplest form includes tools such as forScore that remain specialized in CWN while allowing the display of graphic scores as PDF documents.

Applications like Decibel Score require the composer to create a score using external graphics editing software and import it into the program, which is then converted into a scrolling score with animation features that can be displayed on multiple synced iPad devices [7]. While these tools provide features such as scrolling score display and multi-device synchronisation that are beneficial for performance, their reliance on the use of external graphics editing software for composition may be technically challenging for some.

Other complex software environments such as IanniX [8] and Ossia Score [9] offer solutions for event sequencing and real-time score generation. While these platforms allow dynamic interaction with other software, facilitating synchronization and the triggering of sonic or visual events, their complex graphic and code-based approach can present steep learning curves for some artists or offer too many options for those wishing to focus only on creating a graphic score.

SIGN/e has been developed to fill the gap within this spectrum of graphic notation possibilities by serving as an accessible and easy-to-use platform that works in conjunction with the digital music production capabilities of a Digital Audio Workstation (DAW).

3. FROM GSR TO SIGN/E

Developed between 2018 and 2021 for “Ensemble d’oscillateurs” at the Université de Montréal, Graphic Score Reader (GSR) is a Max for Live plugin designed for real-time scrolling display of graphic scores. The project was directed by Nicolas Bernier and programmed by Evan Montpellier.

GSR supports a variety of image formats. Each Max for Live device can contain up to twelve score files, allowing sequential playback of multiple scores in a concert performance context.

Multiple devices can be loaded simultaneously to project different synchronized versions of the same graphic score (e.g. a clean version for the audience and an annotated version for performers). Score playback can be synchronized with Live’s transport or controlled using automation curves. In 2023, Ensemble d’oscillateurs’ instrumentarium evolved from sine wave oscillators to analog synthesizers and “no-input” mixers. This transformation introduced

new considerations for musical writing and raised unprecedented challenges, paving the way for the creation of SIGN/e.

Like other similar applications, GSR lacked flexibility in score editing as it forced the use of external graphics editing software. Thus, adjusting scores was a cumbersome process: each modification required returning to the graphics software, making corrections, re-exporting, and re-importing the updated score into GSR. Furthermore, the use of external graphics editing programs can be prohibitive for individuals without prior knowledge of such tools. SIGN/e emerged from a growing need to simplify the process by merging writing and reading within the same environment. Furthermore, the integration of SIGN/e into a DAW allows for extensible modular control by using Ableton Live’s parameter mapping functionality, as will be detailed in this article.

4. SIGN/E: SYMBOL-BASED INTERFACE FOR GRAPHIC NOTATION EDITION

4.1. System Overview

SIGN/e operates through four interconnected devices:

- 1) SIGNe-Screen.amxd serves as the central hub for visualization and global control;
- 2) SIGNe-Symbol.amxd manages individual graphic elements and their dynamic properties;
- 3) SIGNe-Text.amxd allows textual notation and animation;
- 4) and SIGNe-Background.amxd enables the addition of a fixed or scrolling image that can serve as a score template or other visual reference created using external graphics software.

This architecture allows composers and performers to integrate visual notation directly into a DAW (Figure 1).

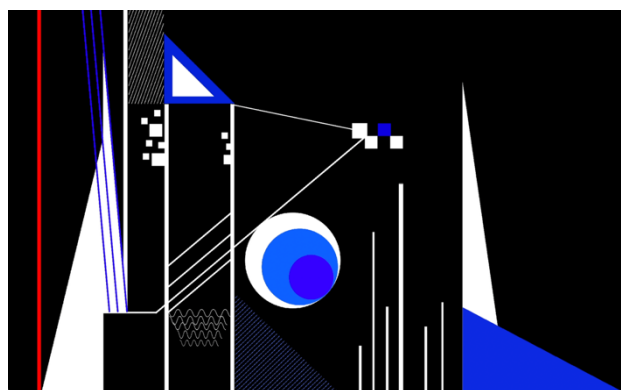


Figure 1. Still from an animated graphic score by Alexandre Sasset-Blouin, entirely created with SIGN/e for Ableton Live.

By combining visual design tools with the automation and sequencing capabilities of Ableton Live, SIGN/e provides virtually endless possibilities. For instance, one could easily map Ableton Live’s random LFO to the height of a symbol so its position would randomly change over time, in synchronicity or asynchrony with Ableton Live’s

timeline; or one could map an envelope follower to the opacity of a symbol so that input from an acoustic instrument would control the appearance and disappearance of a graphic symbol. SIGN/e fosters dynamic interaction between sound and image via its integration with Ableton Live's features, supporting both structured and improvisational approaches.

Its architecture facilitates idiosyncratic forms of notation and performance, particularly in contexts where conventional notation is inadequate for representing and communicating the musical ideas.

4.2. Architecture and Workflow

The SIGNe-Screen.amxd device acts as the host for all SIGNe-Symbol.amxd, SIGNe-Text.amxd and SIGNe-Background.amxd devices added to the session. SIGNe-Screen.amxd generates the video output window where graphic elements are displayed and can be manipulated directly on the screen with the mouse, while providing access to global score parameters via the device's settings window, e.g. background color, quantization, time ruler, etc. SIGNe-Symbol.amxd, SIGNe-Text.amxd and SIGNe-Background.amxd are placed on MIDI tracks. Each device corresponds to a single visual element. The parameters of each symbol are automatable, editable and mappable like any Ableton Live plugin. Depending on preference, the user may choose to use one track for each device or add multiple devices to a track and select the one to automate using Live's Device Chooser dropdown menu.

To use the SIGN/e devices, a SIGNe-Screen.amxd should always be added onto a single MIDI track to display the video window. Subsequent SIGNe-Symbol.amxd devices are added to other tracks, introducing placeholder elements that can be replaced through the Symbol and Background image selector, or the Text editor.

4.3. Global Controls

SIGNe-Screen.amxd (Figure 2) allows users to manage the global settings of the score (Figure 3) via the settings menu. The overlay menu provides access to display, navigation and project parameters. Additional options include among others, Freeze All, which locks all symbols against manual repositioning.

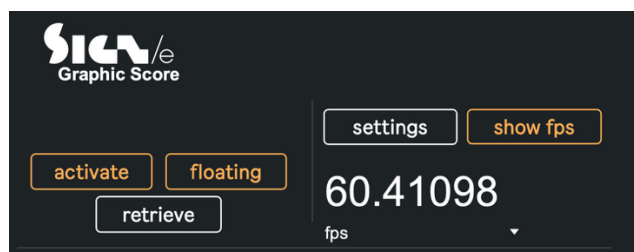


Figure 2. SIGNe-Screen.amxd device

The device setting menu also gives access to a session editable quantized grid, a time ruler showing measure numbers on the bottom of the score, and a scrolling playhead synchronized with Ableton Live's transport. Window

management options include buttons to make the window float over other windows and to restore it to a default position, as well as settings to adjust the background colour and a global framerate indicator.



Figure 3. SIGNe-Screen.amxd settings menu

Playback can follow Ableton Live's transport or be driven by automation curves. Score duration is defined by a numeric value in measures. Driving the score via automation curves allows the score to scroll both backward and forward as the Live set plays. The scrolling position parameter can also be mapped to a controller to facilitate manual control over the movement of the score, making it possible to move through sections as quickly or slowly as one wants. More options are accessible through the device settings interface, and their functions are covered in the user manual.

4.4. Symbol-Level Controls

SIGNe-Symbol.amxd, SIGNe-Text.amxd, and SIGNe-Background.amxd manage the properties and behaviours of individual graphic elements shown in the score window. SIGNe-Symbol.amxd provides visualization of the selected symbol, along with its controls: rotation, position, scale, opacity, layer depth, and its repetition (Figure 4).

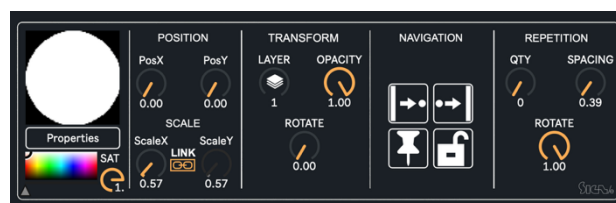


Figure 4. SIGNe-Symbol.amxd device

The device properties allow users to choose a symbol and a pattern while accessing numerous parameters (Figure 5). Colour settings support interpolation between two user-selected colours with automation capability, allowing alternation between the two selected colours via automation or any mappable Ableton Live device (LFO, envelope follower, etc.). The pattern parameters such as tiling and opacity can be adjusted dynamically: for instance, a circular symbol could have a square tiling pattern that would grow bigger over time to emphasize a timbre mutation within the duration of a musical gesture (the gesture symbolised by a circle in this example). The device also offers quantization tools that constrain symbols' movement along the X axis according to rhythmic divisions and the Y axis based on vertical divisions.

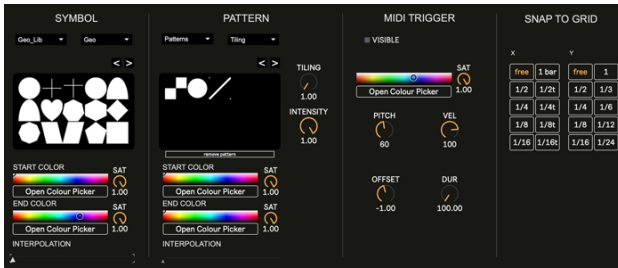


Figure 5. SIGNe-Symbol.amxd device properties

SIGNe-Symbol.amxd and SIGNe-Text.amxd contain MIDI trigger functionality, which extends the interactive potential of each symbol, linking visual elements to sonic events in real time. A MIDI trigger is a point that moves together with a symbol or text object in the graphical score. When a trigger passes the playhead, the corresponding device emits a MIDI note. By default, a MIDI trigger is invisible, but it may optionally be visualized as a vertical line. A trigger always moves in sync with its symbol, but may be displaced horizontally relative to the symbol, allowing manual calibration of the timing between the visual movement of the symbol and its MIDI output. A dedicated part of the device menu allows detailed editing of MIDI trigger parameters, including pitch, velocity, duration, and visual attributes.

Common parameters like rotation, scale, and opacity can also be modified using mouse and keyboard shortcuts for a more efficient workflow.

The SIGNe-Text.amxd device allows you to add text to the score (Figure 6). Text can be used for annotating the score, for passing indications to musicians, to create text scores, or to use text as a purely visual element. The device offers similar position and parameter controls to SIGNe-Symbol.amxd, but includes specific text editing options such as font selection, letter spacing, and the ability to manually adjust the bounding box used for mouse selection.

Finally, the SIGNe-Background.amxd (Figure 7) device offers the option to add a background image that can serve as a score template or other visual reference. The image can be set as fixed or configured to scroll in sync with the other visual elements via automation or by following the session transport.



Figure 6. SIGNe-Text.amxd device

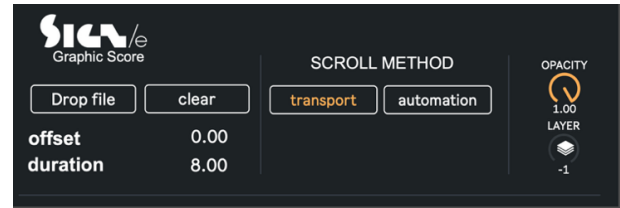


Figure 7. SIGNe-Background.amxd device

4.5. Integration of Custom Graphics

SIGN/e includes a default library containing several categories of useful notation symbols, including simple geometric shapes, arrows, and lines. Additionally, SIGN/e supports importing user-defined graphic elements, which can be added by placing PNG files in the USER directory within the application's LIBRARIES folder. Recommended specifications include a resolution of 300 dpi and dimensions of approximately 500×500 pixels to maintain visual quality when scaling graphical elements. This feature allows composers to personalize their scores and expand the visual vocabulary beyond the default libraries.

5. SOFTWARE FRAMEWORK

5.1. Development Environment and Design Principles

SIGN/e is implemented entirely using only standard objects included in the official Max distribution. While bringing some limitations, this choice was made to ensure long-term maintainability and cross-platform stability, avoiding dependencies on third-party externals. Core operations such as filename parsing and path construction rely on CPU-based Max objects combined with the Live Object Model, which provides access to Ableton Live's internal state.

5.2. Rendering Pipeline and Visual Architecture

SIGN/e utilizes Max's Jitter framework to establish an OpenGL rendering context. Although the score is visualized in two dimensions, the system operates within a 3D space using orthographic projection (a method of rendering a 3D scene in 2D that involves drawing the scene via rays that are perpendicular to the camera plane and parallel to each other, as opposed to having all the rays converge at a single point, as in a pinhole camera or eye) to eliminate perspective distortion. This architecture leverages Jitter's robust feature set, providing native support for geometric primitives, meshes, and procedural text, along with intrinsic parameters for position, rotation, scale, colour, and blend modes. By enclosing these objects in Max for Live devices, the system links Jitter's visual manipulation features directly to Live's automation, undo history, and save systems.

To simulate a strict 2D environment, a custom mouse interaction scheme restricts transformations to a camera-facing plane. A pure 2D fragment shader approach was explicitly discarded; unlike Jitter's object-oriented structure, a fragment shader-based system would require dynamically creating parameters and shader inputs for each new

graphical object via the Max API. This method was rejected to avoid instability and ensure reliable data retention.

All graphic elements are positioned in a plane perpendicular to the 3D camera. The camera moves horizontally in synchronization with Ableton Live's transport, ensuring that spatial coordinates correspond directly to musical time. While standard object properties are controlled via OpenGL, advanced features, such as grid generation, time rulers, and playheads, are implemented using shaders written in Max's Gen language, which compiles to GLSL for efficient GPU execution.

5.3. Parameter Management and Automation

Compositionally relevant parameters are exposed through Max for Live User interface (UI) objects, allowing integration with Ableton Live's automation system. This integration provides several benefits: parameter changes are recorded in real time, stored within Live sessions, and included in the undo stack. Automation data can be authored manually or generated algorithmically, then edited post-recording. For data types unsuitable for direct storage via UI objects, such as strings, the system employs Max's `patrr` preset mechanism, which interfaces with Live's parameter storage framework.

5.4. Modular Architecture and Device Structure

SIGN/e employs a modular architecture that distributes visualization and control across distinct devices. This strategy reflects established practices in 3D content creation and game development where 3D objects are represented as discrete nodes; when selected, the user can edit the parameters of the corresponding object. By allocating one device per symbol, SIGN/e maintains organized automation data and enables scalable resource management. Although this methodology may introduce complexity in extensive projects, it adheres to workflows commonly found in advanced graphics software.

5.5. Symbol Representation and Mesh Design

Symbols are rendered as textures mapped onto geometric meshes stored in FBX format, with UV coordinates configured for camera-facing display. Initially limited to quads, the mesh set was expanded to include circles, triangles, and diamonds to improve mouse interaction and sorting when multiple symbols overlap in space. Textures support alpha channels allowing non-rectangular shapes to be displayed without revealing the mesh shape, and associations between meshes and textures are currently hard-coded.

5.6. Graphical interface interaction

Jitter's native system for mouse and keyboard interaction, particularly through the `jit.gl.handle` object, is not optimized for pseudo-2D objects. The absence of axis-locking for rotation makes it nearly impossible to rotate an object exclusively around the axis perpendicular to the camera, often resulting in off-axis orientations. Furthermore,

mouse selection in `jit.gl.handle` relies on checking the mouse's position in screen coordinates against an object's world-aligned bounding box, which is always a quad aligned to the render context's axes. When two non-square symbols with alpha channels are placed close together, overlapping bounding boxes can cause incorrect object selection, leading to a frustrating user experience.

To address these limitations, a custom interaction system was developed using Jitter's physics engine to generate collision volumes that approximate the actual shape of the meshes used for rendering symbols. This approach significantly improves selection accuracy and introduces support for axis-locked 2D rotation, providing a more intuitive and reliable interaction model for score editing.

5.7. Performance and optimization

Given that SIGN/e is intended for use in real time presentation scenarios (i.e. concerts), performance is a serious consideration. While the current design that relies on a one-to-one relationship between Max for Live devices and graphical objects is convenient from a UX and development perspective, this architecture has limitations in terms of performance. Each instance of a Max for Live device in a Live session requires a certain amount of available memory, with more complex devices requiring more memory. In their current state, each SIGN/e-Symbol device occupies about 50 MB of memory. Since this memory use scales linearly as more devices are added to the set, it imposes a hard limit on the number of symbols that can exist within a given score before performance degrades.

Given the memory limitations of a mid-range laptop, SIGN/e can approximately accommodate 300 SIGNe-Symbol.amxd devices. Hence, the users must work within these limits, e.g., use a limited number of symbols per session; use the fixed background image feature to reduce RAM usage; or separate a score into multiple sessions. These limitations vary according to the available RAM of the computer used. We are currently exploring several optimization strategies for the next development stages.

6. CONCLUSIONS

SIGN/e is designed for artists (composers, musicians, designers) who want to create graphical music scores within the Ableton Live architecture using a simple and accessible interface. The simplicity and accessibility of the proposed tools aim to make graphical notation available to creators who are not familiar with often complex and costly graphics editing applications such as Adobe Illustrator or Affinity Designer.

While we understand that SIGN/e's exclusive integration with Ableton Live could be seen as a drawback for non-Ableton Live users, we believe that integration with a widely-used DAW is key to contribute to the democratization of graphic scores. By embedding scoring tools into a program that a critical mass of musicians uses every day, SIGN/e should appeal to a community that may not typically write scores or possess a background in contemporary notation. This immediacy removes the friction of switching contexts; users who might never seek out

standalone scoring software may be compelled to experiment with graphical notation simply because it is available within their native production environment.

Furthermore, the modular, device-based approach chosen for SIGN/e provides the opportunity to add more devices in the coming years and allows users to create custom ones that can be integrated into the SIGN/e architecture, thus making the application customizable and virtually endlessly expandable to meet specific needs.

By providing an accessible and flexible environment for creating animated graphic scores, the system bridges the gap between conventional notation tools and the creative demands of graphics-oriented compositions. Its architecture not only simplifies the process of score creation but also allows dynamic interaction between sound and image, supporting both written and improvisational approaches.

Future developments will focus on extending the device library, enabling user-defined meshes, and improving interaction models to further enhance customization and usability. Through these iterations, SIGN/e aims to contribute to the evolution of performative notation and to promote new forms of collaborative and experimental music-making.

Acknowledgments

This article is part of the Performative Notation project directed by Nicolas Bernier and funded by Fonds de recherche du Québec – Société et culture (FRQSC), a research project held at Laboratoire formes • ondes (LFO) at Université de Montréal between 2024 and 2027. Pierre-Luc Lecours would like to thank the Social Sciences and Humanities Research Council of Canada (SSHRC) for funding his doctoral research between 2021 and 2024 and the Centre for Interdisciplinary Research in Music, Media, and Technology (CIRMMT).

7. REFERENCES

- 1.N. Bernier, “Observations on Performing Sine Waves with an Oscillator Ensemble,” *Leonardo*, vol. 55, no. 2, pp. 161–165, 2022.
- 2.P.-L. Lecours and N. Bernier, “Sound Synthesis Notation Applied to Performance: Two Case Studies,” in *Proc. of the Int. Conf. on Technologies for Music Notation and Representation – TENOR’24*, Zurich University of the Arts, 2024, pp. 142–149.
- 3.J. Bresson, C. Agon, and G. Assayag, “OpenMusic: Visual programming environment for music composition, analysis and research,” in *Proc. of the 19th ACM Int. Conf. on Multimedia*, Scottsdale, Arizona, 2011, pp. 743–746.
- 4.H.-W. Nienhuys and J. Nieuwenhuizen, “LilyPond, a system for automated music engraving,” in *Proc. of the XIV Colloquium on Musical Informatics (XIV CIM 2003)*, vol. 1, 2003, pp. 167–171.
- 5.G. Hajdu and N. Didkovsky, “MaxScore: Recent developments,” in *Proc. of the Int. Conf. on Technologies for Music Notation and Representation – TENOR*, vol. 18, 2018, pp. 138–146.
- 6.A. Agostini and D. Ghisi, “A max library for musical notation and computer-aided composition,” *Computer Music Journal*, vol. 39, no. 2, pp. 11–27, 2015.
- 7.C. Hope, L. Vickery, A. Wyatt, and S. James, “The DECIBEL Scoreplayer—A Digital Tool for Reading Graphic Notation,” in *Proc. of the First Int. Conf. on Technologies for Music Notation and Representation – TENOR’15*, 2015, pp. 58–69.
- 8.T. Coduys and G. Ferry, “Iannix aesthetical/symbolic visualisations for hypermedia composition,” in *Journées d’Informatique Musicale*, 2004.
- 9.J.-M. Celerier, P. Baltazar, C. Bossut, N. Vuaille, J.-M. Couturier, and M. Desainte-Catherine, “OSSIA: towards a unified interface for scoring time and interaction,” in *Proc. of the First Int. Conf. on Technologies for Music Notation and Representation – TENOR’15*, 2015, pp. 81–90.