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 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 nirgun bhajans: spiritual songs addressed to nameless, non-anthropomorphic entities. In this piece, singers produce

[1] Non-linear and open notational strategies have a long history in 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 comprovisation 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.

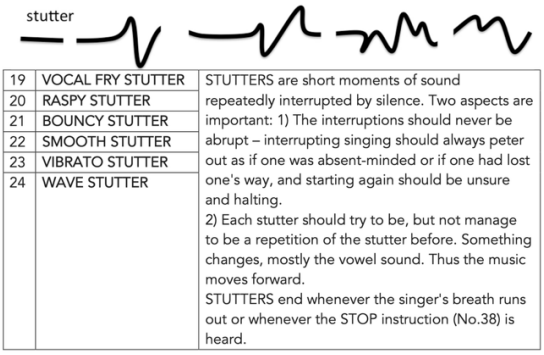


Figure 1. Example of a vocal instruction legend.

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.

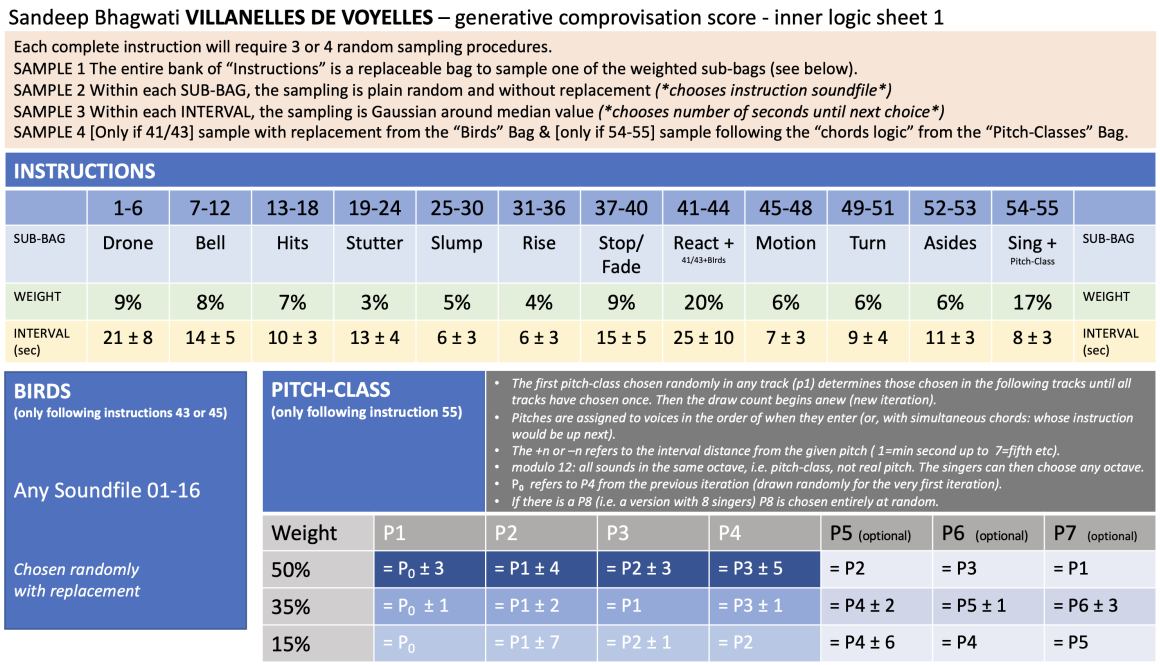
For generative purposes, all instructions are grouped into thematic sub-bags (e.g. drones, bells, hits, motion, turning, singing with pitch-class logic). Each sub-bag is assigned a probabilistic weight that determines how frequently instructions from that category are selected over time

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 comprovisation 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_0 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 remain 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, reorienting, 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 soundfiles 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-

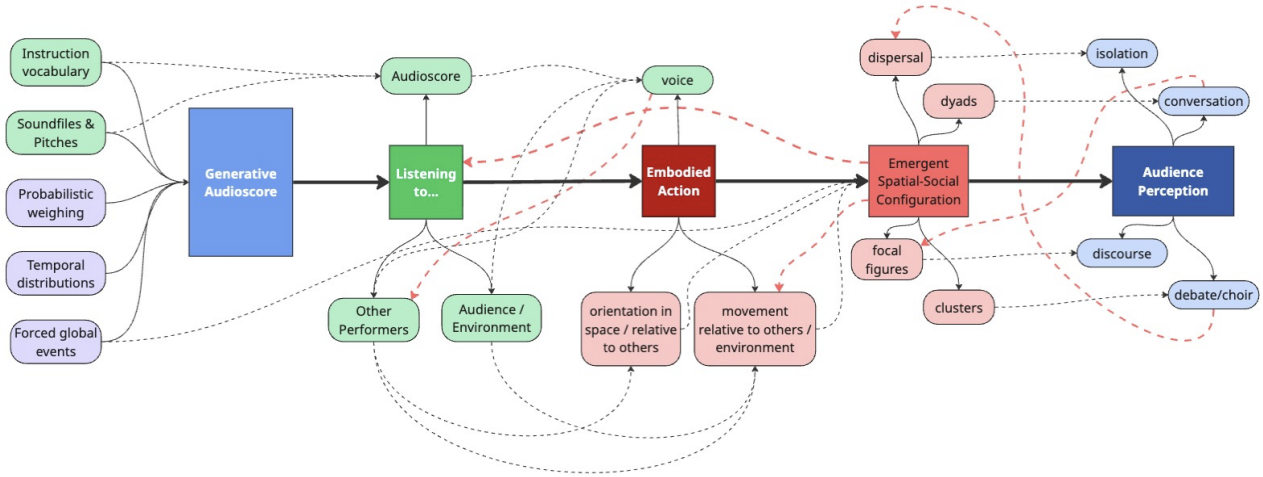


Figure 5. Visualization of the generative audioscore of Villanelles de Voyelles as a multi-layered process in which listening, rather than gesture or movement, functions as the primary motor of musical, spatial, and social action. Although organised as a left-to-right flowchart, the diagram includes 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 2022 [3] . 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.

[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.

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.

[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