LINEAR (LIVE-GENERATED INTERFACE AND NOTATION ENVIRONMENT IN AUGMENTED REALITY)

Giovanni Santini
Hong Kong Baptist University
Kowloon Tong
Hong Kong
info@giovannisantini.com

ABSTRACT

Recent developments in Augmented Reality (AR) technology are opening up new modes of representation and interaction with virtual objects; at the same time, increase in processing power of portable devices is enabling a wide diffusion of applications until recently usable only in very specific situations (like motion-capture labs).

This study aims to describe an AR environment created for musical performance: LINEAR (Live-generated Interface and Notation Environment in Augmented Reality), where the author explored some perspectives made possible by the current state of AR technology applied to music.

In LINEAR, one dedicated performer using an AR iPhone app, can create virtual objects (rendered in real-time and superimposed to the real environment) according to the movement of the device; they are used both as virtual interfaces for electronics (sending OSC message to Max/MSP on a computer) and as forms of live-generated graphic notation. LINEAR allows, with some limitations, the representation of gestural movements with an exact 3-D placement in space: we can now have an analogic notation of gestures, rather than a symbolic one. For the iPhone performer, the act of notation corresponds to the notated act.

The resulting representations can be also approached as graphic animated notation by other performers (the iPhone screen is mirrored to a projector).

The multiple perspectives on the notation and the possibilities of interaction with virtual bodies allow a high level of flexibility, while introducing some almost unprecedented resources and foreseeing a very rich scenario.

1. INTRODUCTION

The idea of LINEAR came from a simple observation: no kind of existing musical notation can really represent a gesture. Even if a specific movement can be described through some kind of symbol or graphic representation, its trajectory can never be fixed in space. The importance of gestural notation in musical scores has been increasing with the overwhelming exploitation of Extended Techniques and the implementation of choreographies inside the compositional process, sometimes even detached from the need for a resulting sound.

The current state of AR technology allows the live generation of virtual entities that can keep track of the trajectory of one gesture in 3-D space. Those bodies can then be linked to arbitrary functions and data, thus constituting virtual interfaces. Furthermore, the trajectories can be interpreted as live-generated graphic notation.

The implementation of these possibilities inside LINEAR is aimed at developing open forms with an open instrumentation (including live signal processing); the real-time generation of the score and the possibility of change in perspective (thanks to AR technology) create a lively compositional ecosystem: the score is not pre-composed by the composer; instead, every performer has the possibility to intervene in real time on the notation, while interpreting it. This way, every player can influence the other ones' behavior. As explained in section 3, this is particularly true for the iPhone performer, who has the highest level of control on the notation.

The project described in this paper was not conceived to explore all the possibilities offered by AR applied to music. It is, instead, a work in progress, where some preliminary ideas are realized, revealing limits both in the still new, fresh and basically unexperimented practice of AR based musical performance [1] and in the technology itself.

2. BACKGROUND

The development of LINEAR is based on a very new evolution in technology and therefore the author could not rely on numerous similar experiences realized before. However, the artistic and technical panorama providing a background for this work is quite vast. In the next paragraphs, the different aspects of such scenario will be introduced.

2.1 Graphic notation on paper

Since the 50s (and in isolated cases even before [2]) musical notation has been pushed beyond a pitch-rhythm representation (as in Common Western Music Notation), in favour of an enormous amount of experimentations, depending on

[1] Visual augmentation of live music performances (as in [1]) cannot be assimilated to the intended outcomes of this project, since that kind of visual augmentation does not have direct consequences on sound and gestural behavior.

[2] E.g., L. Russolo, *Risveglio di una città, 2014*.

different aesthetical purposes, authors, environments and historical periods.

For this reason, a brief categorization of graphic scores or forms of graphic notation can never be really exhaustive or precise. However, we could roughly divide the use of graphic notation in five main categories:

- graphics are linked or linkable to specific parameters (for example durations or dynamics), even if in a "non-conventional" context (as in Cage's *Variation II*, 1961);
- electronic music notation (an example could be Stockhausen's *Studie II* (1954); however, depending on the aim of notational process and kind of composition, the notation could vary enormously);
- graphics are used to obtain some kind of intuitive reaction and forms of free association; they may result from particular re-combinations of traditional staves (as many pieces in Crumb's *Makrokosmos*, or in Bussotti's *AutoTono*, 1978);
- all the possible structures/trajectories of the work are resumed in one map (as in Kourliandsky's cycle of *Maps of non-existent cities*, 2012) or in a rhizomatic [3] representation (as in Haubenstock-Ramati's *Konstellationen*, 1976);
- graphics are used (often even in combination with traditional notation) in order to add indications of specific gestures/actions basically (although not necessarily) aimed at producing sound (as in Lachenmann's *Gran Torso*, 1971 or Laporte's *Dégonflement, 1978* or Yiran Zhao's Dirigentenquartett *Verwickelte Synästhesie)* (2013).

As showed later, the graphic notation generated in LINEAR may be referred to the last two categories.

2.2 Real-time scores and animated notation

As soon as the technology allowed it, the gain of a temporal dimension inside a score (i.e., time is not just codified on the x-axis of the paper, but really "passes" and modifies what the performer sees) became another way to push notation beyond its "traditional" boundaries.

Essentially, real-time scores make use of some forms of animated notation (i.e., graphic animation is implied, producing scrolling, permutation, transformative or generative scores) [3].

According to Freeman, real-time scores can be placed "in the context of algorithmic and computer-assisted composition and also within the aesthetic framework of open-form composition" [4]. Usually, they require a constant sight-reading since the performer cannot wholly foresee the following musical events.

For example, Gerhard Winkler's *KOMA* (1995) makes use of live-generated scores, visualized on a computer, where shapes related to micro-glissandos and dynamics are

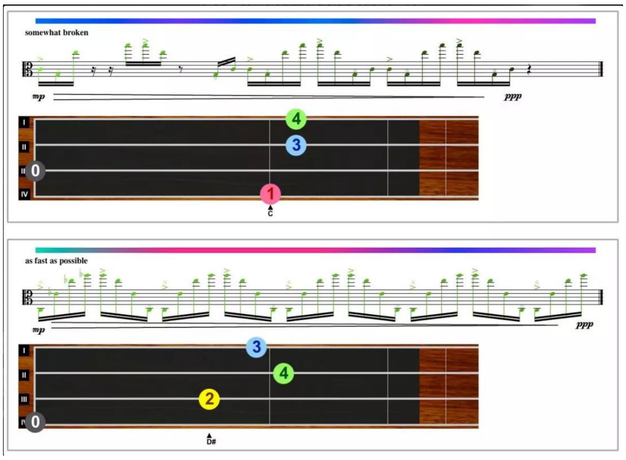

Figure 1. Extract from S. Shafer's *Terraformation*. The color-gradient line on top indicates bow contact position, "fret" notation indicates fingerings and coloured circles left-hand pressure over the strings.

continually moving in real-time, according to principles of real-time generation [5].

In Shafer's *Terraformation* (2017) chords to be performed on a viola are created during the performance (following specific rules set by a decisional algorithm) and translated, in real-time, into an action-based notation comprising three different layers (common notation, "fret" notation and two sets of color-gradient notation, see Figure 1).

The *Decibel Scoreplayer* is a tool for real-time scores, allowing the network-synchronized scrolling of graphic scores ([6], [7]).

In some cases, it is used for regulating the real-time changing transparency of different superimposed images forming a score; such changes in transparency allow the visualization of different trajectories/possibilities inside a rhizomatic score (as in Vickery's composition *trash vortex* [7], [2]).

Some compositions reveal a strong orientation towards an advanced use of graphics and animations, almost transcending the concept of score in favor of the idea of dramaturgy. For instance, in P. Turowski's *Genni* (2018, Figure 2), the score may be seen as the staging of a plot with geometric figures as characters. This piece also shows the expansion of animated notation to the third spatial dimension. In the next paragraph, 3-D notation will be presented more in detail.

2.3 3-D and VR scores

The possibility to access the third dimension in image rendering in real-time, made possible by the increase in processing power of computers and the diffusion of programming frameworks for real-time 3-D rendering (as *Jitter* and *Processing*) can be considered, in the opinion of the author, a real turning point in musical notation.

Kim-Boyle, in [8], presents two compositions using 3-D notation. In *16:16* for piano, in particular, the score is animated and nodes inside 3-D space are mapped to different

[3] For the concept of rhizomatic musical notation, see [2].

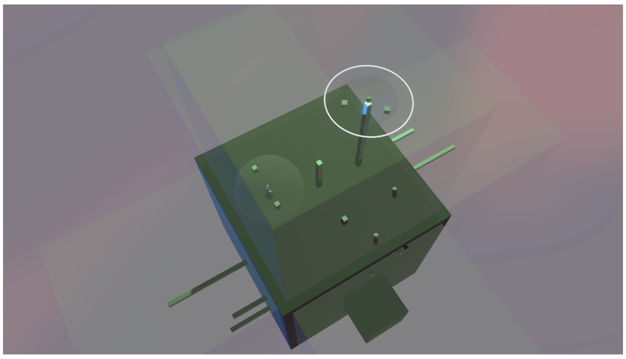

Figure 2. Screen-capture from P. Turowski's *Genni*.

pitches and kinds of piano strings preparation according to position and color. The author also proposes the visualization of the 3-D score by using red-cyan glasses; such an adjustment allows a true perspective vision of 3-D notation for more than one person at a time without the need of VR setups.

Another interesting form of 3-D notation, for playing drums, can be found in [9]: the score allows the representation of different layers of information about the same drum pattern, depending on the point of view on the 3-D structure obtained from pattern analysis; the author also shows 3-D printed scores obtained by those models. Finally, he introduces the use of VR for immersive visualization of 3-D models.

The first ideas about musical visualization in VR can be traced back to 2001 [10], with the proposition of a virtual representation of musical structures derived from form analysis in a VR environment; however, only recently some real experiences have being developed.

In *SpectraScore* [11] (Figure 3), elements visualized in a 3-D VR environment (rendered in real-time) transmit image data to Max/MSP for audio synthesis. Therefore, the sound environment changes depending on visual data extracted from the observed objects.

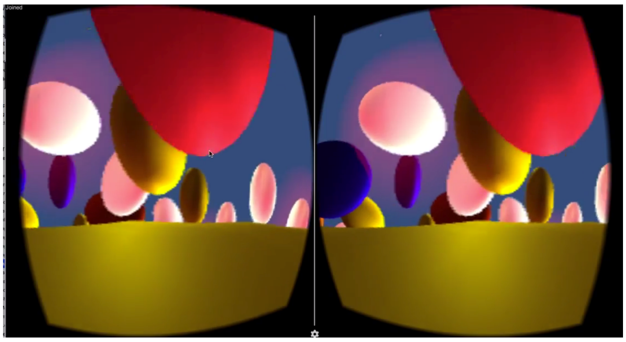

Figure 3. Screen-capture showing the stereoscopic vision for a session with SpectraScore VR [Available: https://www.youtube.com/watch?v=LK6rQFAmPDE].

In A. Brandon's *Hidden Motive* (2018), a graphic score is generated live by the composer (who may also be in another part of the world, sending it through wi-fi) and transmitted to a mobile device mounted on a VR headset (Figure 4). The score is also mirrored to a projector.

However, the use of VR is not necessarily aimed at visualizing 3D scores. In *[P.O.V.]* (2017) for saxophonist, VR glasses, electronics and video mapping by Oscar Escudero Romero, the performer uses VR glasses for visualizing a 2-D scoreplayer and some short animations used as markers for some musical details (like repetitions). The use of VR, in that case, is necessary because of the particular nature of the piece: lights should be turned off in order to deliver good quality projections; this solution lightens up another potential use of VR: scores can be visualized even in the absence of light.

All the experiences above, from graphic to VR scores, extend resources and aims of notation far beyond the Common Western Musical Notation. If a trend can be traced in the presented research progress, it consists in a process towards forms of 4-D representation (notation in the space-time continuum) and interactivity. Last developments in AR technology could constitute the most advanced peak in that direction.

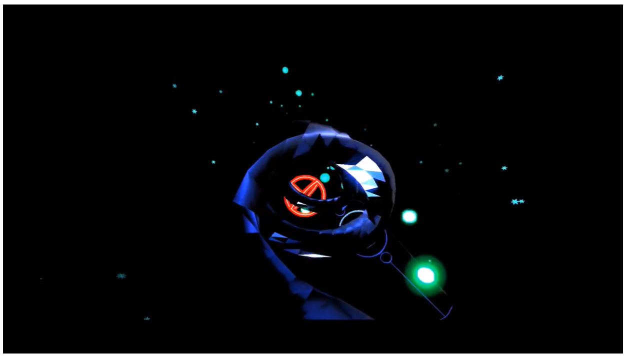

Figure 4. Screen-capture from Amy Bandon's *Hidden Motive*.

2.4 Augmented Reality

Augmented Reality (AR) is a term coined in the 1990s by Tom Caudell [12], for a technology born in the 1960s ([13], in [12]). AR allows the vision of virtual objects (with a precise position in space) superimposed onto the real environment; those "holograms" are made visible through the use of portable devices such as smartphones or tablets (one example is the famous *Pokmon Go!* developed by *Niantic* [4]). In different setups, virtual objects are visualized through screens or projectors connected to a computer (as in [14]). The use of head-mounted see-through devices (such as HoloLens [5]) may be referred to AR, although it is usually inscribed inside the Mixed Reality (MR) [6] field.

Another essential feature of AR consists in the real-time interaction with those items: 3-D virtual objects have a precise position in space and can be looked at from different perspectives. They can be manipulated, with some limits, according to their shape and position in space.

[4] https://www.pokemongo.com

[5] Microsoft HoloLens: https://www.microsoft.com/en-us/hololens

[6] MR is that technology that allows the representation of virtual objects inside the real environment through the use of specifically designed see-through head-mounted devices.

2.5 AR and music

The first experiments in the application of AR to music can be traced back to 2000 with the *Augmented Groove* [15]. The study proposed a marker-based [7] AR sequencer, where different people could insert or remove MIDI tracks by adding or removing cards from a table. Users wearing head-mounted displays (not see-through) could see virtual images rendered on top of the cards. Similar techniques were used by the same authors in 2001 [16], 2003 [17] and in 2007 [18] (implementing also voice and gesture recognition for interaction). A similar experience, yet more evolved and expanded to the simultaneous use of more than one setup (a sort of orchestra of marker-based AR instruments) can be found in [19].

The literature started to grow especially during the last years, following the increased technological possibilities and the continuously spreading interest in the market. Applications developed so far seem constrained, at least in most of the cases, to the imitation of already existing interfaces (as the holographic interface for Behringer DeepMind [8]) or to an aid for improving learning in already existing practices on traditional instruments (e.g., [20] and [21] for guitar, [22] and [23] for piano).

For example, in [21] AR technology was used as a support for studying different songs for guitar: virtual fingers were projected on the frets (visualized on a screen connected to a computer), in order to indicate positions for specific chords. The virtual fingering positions were changed according to the exact timing of the selected song.

Augmented Piano Roll [22] and *Pianolens* [23] (Figure 3) [9] have many similarities. While there are some differences in implementation (use of projectors or HoloLens), the functioning is almost the same. Some colored blocks, whose width corresponds to the keys' one and whose length is proportional to the duration of the note to play, are rolling towards the performer. When they come across a specific line (which is the indicator for "now"), the pianist has to press the corresponding key and keep it pressed until the end of the block.These systems also provide feedback on right and wrong notes and rhythms.

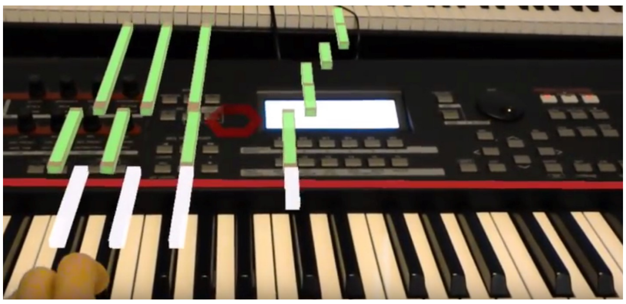

Figure 5. PianoLens demonstration.

[7] Marker-based AR renders virtual objects according to image recognition of specific markers. Each marker is related to one precise virtual objects. Usually, those markers are drawn on cards and the 3-D model is rendered on top of them.

[8] *Behringer DeepMind 12 Augmented Reality Launch*: https://www.youtube.com/watch?v=-9MTlsA-wi4

[9] https://www.youtube.com/watch?v=5TExa2L1rOM

Other studies have been focusing on the exploration of potentialities for real performance.

An interesting use of virtual objects in AR as control interfaces is explored in [14], describing an environment where controllers are visualized in the real world through the use of projectors. Such interfaces can be used thanks to the spatial tracking provided by the use of depth and RGB cameras.

GLASSTRA [24] allows the conductor of a laptop orchestra to visualize in real-time the status of the orchestra on Google Glasses.

An app created by the media artist Zach Liebermann [10] permits the recording of sound while generating a 3-D sound-wave representation. Different visual chunks of the virtual object are linked to correspondent audio chunks in the recorded sound (hence back-and-forth movement along the drawing corresponds to back-and-forth playback of the sound file). The 3-D virtual representation also becomes a 3-D virtual interface for playing back the recording.

The HoloLens AR interface for Behringer's *DeepMind 12* provides AR [11] controllers for the synthesizer to be used with bare hands.

All the experiences above present some form of interactivity and imply mostly real-time information delivery. As shown in the next section, LINEAR permits interaction with virtual objects, while allowing a new form of notation that it is not conceived as an aid for learning a previously existing score but as an autonomous musical representation.

3. LINEAR

3.1 Introduction

LINEAR is an environment designed for new forms of notation and new interfaces in Augmented Reality.

It is composed of the combination of different devices. Its core consists in an AR app for iPhone, developed by the author. It communicates with Max/MSP through OSC (Open Sound Control) connection. The iPhone's screen is mirrored to a streaming box connected to a projector. A dedicated router allows wi-fi connection between the devices.

LINEAR is conceived for the development of open forms in an electroacoustic context with an open instrumentation. An ensemble using LINEAR should include the following figures:

- one iPhone performer (using the AR app);
- one laptop player, controlling some parameters and presets of real-time DSP (Digital Signal Processing);
- at least one instrumentalist playing an acoustic instrument (with live processing).

Before presenting functioning and aims of LINEAR, a preliminary technical introduction is necessary.

[10] *AR app - recording sound in space and playing back by moving through it*: https://www.youtube.com/watch?v=ET2CKUqdPCo

[11] See note 8.

3.2 Technical framework

The software on iPhone is based on ARKit, the framework released by Apple in 2017 for developing Augmented Reality applications on iOs devices.

The AR technology provided by ARKit consists essentially in rendering virtual objects over the rear camera input (thus blending these bodies with the real environment); virtual entities have a precise position in space and, at each video frame (the usual frame-rate is 60 frames per second), they are rendered on the iPhone screen according to the perspective of the observer (more specifically, the device calculates its own position and orientation and therefore derives the observer's perspective). This rendering procedure gives the illusion of a precise positioning of virtual projections in 3-D space. Thus, the device's positional tracking is one of the core features of AR.

According to the Apple Developer Documentation [12] , positional tracking (fundamental for correct rendering of 3-D images) is performed through a Visual Inertial Odometry (VIO) algorithm. It is based on two different data sources: CoreMotion (the Application Programming Interface, or API, that delivers combined data coming from gyroscope, measuring orientation, and accelerometer, measuring acceleration) and the iPhone camera. Feature points are extracted from visual data contained in each video frame captured by the camera; they are compared to the contiguous video frames' feature points for understanding spatial movement. At the final stage, feature point analysis is combined with CoreMotion data to provide a stable [13] (as much as possible) positional tracking.

Rendering can be performed in three different frameworks: SpriteKit (2-D rendering, not suitable for the purposes of LINEAR), SceneKit (3-D rendering), and Metal (3-D custom rendering: the most advanced and efficient but requiring low-level programming).

The AR app for LINEAR is developed in Swift using ARKit and SceneKit. For the purposes of this app, one major advantage of SceneKit over Metal lies in the possibility to instantiate an object by coding only its position in space, its shape and its texture. The framework handles automatically the rendering pipeline and the use of projection matrices for providing a convincing spatial perspective.

The library *SwiftOSC* by Devin Roth [14] is included to handle OSC (Open Sound Control) messages.

3.3 The AR app on iPhone

3.3.1 Startup

On start, the app presents the camera view (i.e., the normal input of the device's rear camera). The screen orientation is locked on landscape mode.

A small green sphere is instantiated 50 cm in front of the camera, marking the center of the point of view. At each frame, the sphere's position is updated according to the device's position and orientation, so that it appears always in the center of the screen (i.e., the center of the camera view).

The app has two main functionalities:

- Creating virtual objects (divided into four categories) linked to stored information (including the name of each object, in order to recall precisely the memorized data);
- Sending to and receiving different sets of messages from Max/MSP via OSC according to specific events.

3.3.2 Creation of virtual objects - first three categories

Virtual objects are divided into four categories, each one with a different particle system [15] attached (linked to different colors: yellow, blue, red and dark violet. Figures 6 - 7).

The creation of objects of the first three categories (yellow, blue and red particle effects) is enabled when the iPhone performer presses the lower part of the screen. The device behaves then like a brush, painting virtual entities in space according to the trajectory during the drawing action. The resulting lines are formed by a succession of small, sphere-like virtual bodies (surrounded by a particle system) aligned along one trajectory.

The body category is chosen according to the speed of the device (depending on specific thresholds).

Every time a new body is created, the software gives it a name and links it to the desired set of information (changes may occur according to different setups and instrumentations for different performances).

When the iPhone performer is not pressing the lower part of the screen, no virtual body is created and the application detects collisions [16] between the green sphere and the painted trajectories.

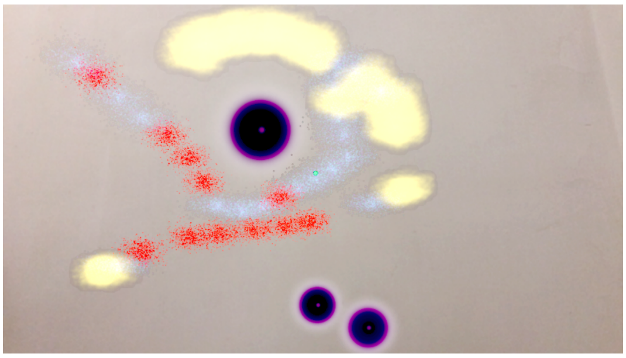

Figure 6. Screen-capture of the iPhone screen running the AR app for LINEAR. One possible graphic result.

[12] Apple ARKit: https://developer.apple.com/arkit/

[13] As better explained in Paragraph 4, the functioning of image data analysis is crucial, since the positional tracking is not stable in case of environments with scarce visual complexity.

[14] devinroth/SwiftOSC: https://github.com/devinroth/SwiftOSC

[15] A particle system (or particle effect) is a graphics effect making use of numerous copies of a small virtual object (particle); each particle can have different movements and behavior. However, the overall impression gives the idea of a single, lively body.

[16] Each object has a "physics body", used for detecting virtual collisions, attached to it. Every time the green sphere marking the point of view collides with one virtual object, the data linked to that object are sent to Max/MSP.

3.3.3 *Creation of virtual objects - fourth category*

The objects of the fourth category (dark violet) are generated according to OSC messages received from a Max/MSP patch running on a laptop. Those OSC messages contain the 3D coordinates of the position of the object to be instantiated; those coordinates are derived from a set of three sound descriptors (e.g., spectral centroid, spectral spread, spectral magnitude) referred to the analysis of the input signal of the Max/MSP patch; the sound produced by one or more instrumentalists participating to the performance is the audio input. The bodies of this category are not generated continuously; their instantiation is triggered by an envelope follower. Additionally, the laptop player can activate/deactivate this functionality.

The iPhone performer can delete every object created in the scene by tapping the highest portion of the screen. All the virtual objects are released from memory, particle systems associated disappear and the data related to the previously created virtual bodies are reinitialized. This function makes it possible to draw new sets of trajectories without preserving the old ones. Such processes are similar to the starting point of a new section in a composition using traditional notation.

3.3.4 *VR mode*

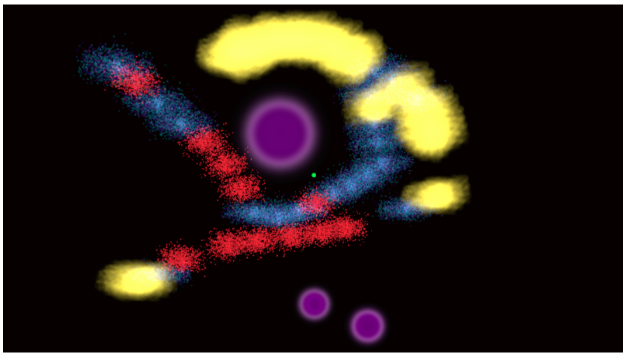

Figure 7. The combination of Figure 6 in VR mode.

The laptop player can trigger a VR mode, excluding the rendering of the camera input and leaving only the virtual bodies. The background can be in any color chosen by the laptop player. This functionality can be used freely throughout the performance. During the VR mode the positional tracking is still functioning, allowing a correct visualization of virtual bodies according to different perspectives.

3.3.5 *OSC communication with Max/MSP*

Information exchanged via OSC is of three kinds:

- data related to virtual objects, sent from the iPhone when bodies are instantiated or whenever a collision between the green sphere and a virtual body is detected;
- speed data sent out at each video frame;
- data related to sound descriptors applied to the input of the Max/MSP patch (acoustic instruments), sent from the laptop to the iPhone.

Figure 8. Rehearsal (F. Teopini iPhone, L. Y. W. Angus flute).

3.4 Production of sound in Max/MSP

Data sent from the iPhone on body collisions are used to play single samples from different libraries (linked respectively to the first three categories). Objects of the fourth category (violet) are linked to the sound they are generated by (the input from acoustic instruments analyzed and sent to the iPhone). Each body, once created, is related to a single sample; therefore, each trajectory drawn by the iPhone performer has a precise sounding identity and can be played in every direction (depending on how the point of view is moved: backward, forward, in small chunks). Objects of the fourth category are discrete points in space (they are tendentially not positioned along trajectories). They break the general continuity of the notation.

The iPhone performer can walk around or across virtual bodies, thus changing the perspective on (and somehow reshaping) the AR interface and the sounding gestures.

The Max/MSP patch also provides live DSP for all the instruments involved in a performance[17].

Speed data are sent at each video frame from the iPhone to Max/MSP and used to regulate different parameters (such as loudness, DSP presets or parameters' values).

3.5 The laptop player

This performer handles volumes, presets, overall balance and spatialization. Additionally, he/she can also choose what the iPhone speed is going to control.

As explained before, Max/MSP sends messages to the iPhone app, in order to create virtual objects according to sound descriptors. This functionality is triggered by the laptop player and can be interrupted by him/her at any moment.

[17] The Max/MSP patch is modular and allows fast implementation of different techniques according to the needs of different performances.

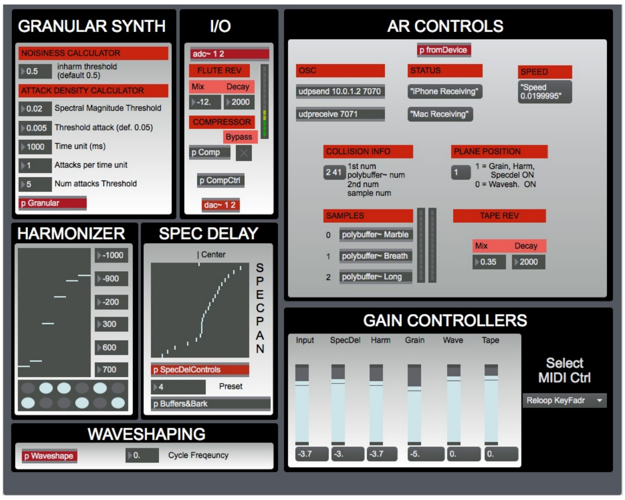

Figure 9. Max/MSP Patch for the laptop player LINEAR (may change for different performances and setups).

Figure 9 shows the UI (User Interface) used by the laptop player for a performance including electric guitar and cello. Techniques and layout may vary depending on the context.

3.6 The perspective of the iPhone performer: graphic gestural 3D notation and virtual tangible scores

The iPhone performer creates the highest number of virtual sounding bodies during a performance, each with a precise placement in space and each linked to a precise sound sample. Consequently, a specific result derives from a specific movement, and that movement is represented by a specific trajectory drawn thanks to the created virtual bodies.

The notation indicates a precise gestural behavior for the iPhone performer: it describes what gesture he/she has to perform in order to obtain a specific result. However, the notation is conceived to leave some decisional freedom to the interpreter, as it does not indicate how fast or how continuously the trajectory should be followed. Furthermore, the performer's movements are not necessarily constrained to the painted lines.

3-D drawings are, at the same time, a control interface for sample libraries. The "physical" interaction between the green sphere marking the center of the point of view and the other virtual bodies generates sounds (through the Max/
MSP patch). In short, from the perspective of the iPhone performer, virtual bodies have two different functions: they bring information about movements for generating sound and they are the "generators" of that sound.

Such a co-presence of notation and sound generator/ control interface in the same virtual objects, induces us to consider the existence of a new typology of scores which could be called (quite oxymoronically) *virtual tangible scores*, as a particular case of tangible scores (defined as "graphical scores [...] physically incorporated in the form of the instrument" [25]).

3.7 The perspective of the other players

3.7.1 Graphic animated notation

The other players (laptop performer and instrumentalists) cannot interpret the drawings the same way as the iPhone performer does. They cannot interact directly with virtual objects.

For them, those trajectories are part of a real-time animated [18] score that does not have immediate gestural implications . The score they read is intended as a means to convey creative energies during the performance. As Fischer writes:

> "An animated notation is an invitation for composers and performers to start their own so-called mapping process. They need to connect or map visual attributes with sonic attributes. In staff notation the mapping by composer and performer are basically congruent. In animated notation the mapping process is done individually, first by the composer and then by the performer." [26, p. 35]

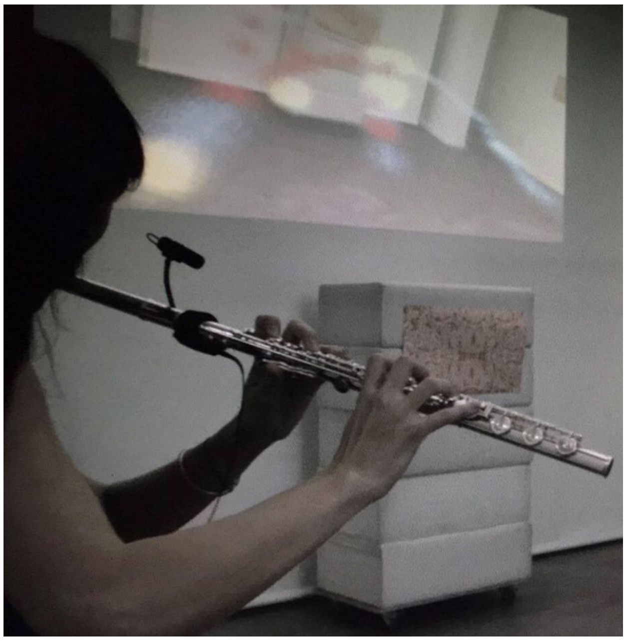

Figure 10. P. Pakiela reading the graphic notation from the projected image.

Such a continuous re-mapping process implies a particular form of creativity that other types of notation intrinsically exclude (e.g., Common Western Notation). The score is not written once and fixed. This kind of notation, derived from graphic scores developed since the 1950s, redefines the idea of composition in terms of continuous creative effort exploited by a group of performers; in LINEAR, in particular, the group operates exclusively in an *in-time* dimension, excluding any *out-of-time* structural planning [19] .

[18] The animation derives from the continuous movements of the iPhone performer, who is constantly changing the perspective on the AR shapes.
[19] Though with a strong simplification, the idea of the in-time/out-of-time dichotomy is derived from Xenakis [27].

Codification strategies applied so far by the author of this article require the identification of different body categories (particle system colors) with generic sound qualities (e.g., yellow = moaning sounds, red = inharmonic-distorted sounds, blue = lively high-pitched sounds, violet = static-low sounds). Instrumental techniques used for the realization of the score are discussed before rehearsal or performance, but there is no pre-decided path to follow. Performers are asked to “read” a single trajectory until they reach a point of conjunction with other trajectories. At that point, they can jump to another tone.

The definition of more refined strategies for the interpretation of the animated score is an open process: the interaction between the iPhone player and the score creates a high number of unpredictable situations, hard to codify in advance. However, some guidelines seem to emerge. The possibility of a change in perspective performed by the iPhone player is probably the most powerful resource: getting closer to (thus zooming-in on) a specific point inside a trajectory has wide repercussions. For instance, getting extremely close to one virtual body would fill the entire screen with the color and particle effect of that body, thus creating a sense of totality of the sound quality related to it. It is also instinctively translated into a *ff* (at least, according to the performers that played in LINEAR). On the contrary, finding a point of view that excludes almost every virtual object, except for some small, far bodies, could be interpreted as a perforated and quiet sound texture.

3.7.2 Notational feedback

The instrumentalists and the laptop player can partially modify the score.

The laptop player can trigger the VR mode: when it is active, the real environment is not rendered and is substituted by a plain color background. In this case, performers can concentrate only on the virtual score. According to the performers the author worked with, this functionality somehow changes the re-mapping process and, in general, makes it easier for the players to concentrate only on the score reading. However, the AR mode is considered the main one. Motivations for this are presented in 3.8. Another functionality is triggered by the laptop player: the creation of virtual bodies of the fourth category.

When instrumentalists play (interpreting the notation on screen), they generate notation (and virtual interfaces), as virtual bodies are instantiated according to sound analysis. Such a process produces a phenomenon that could be called *notational feedback*: the notation is created as an effect of its interpretation. The concept of notation can be pushed to some unknown boundaries, where the ideas of authorship, composer, form and improvisation can be seen under a new, slightly different light.

3.8 Compositional ecosystem

Going back to the definition of animated score by Fischer, we can suggest that, in LINEAR, AR scores go somehow a step beyond. The process of mapping is not done “first by the composer and then by the performer”. It is rather reconstructed in real-time by the whole ecosystem formed by all the performers. The role of the “composer” is limited to the predisposition of the conditions for the ecosystem to be formed (software development, proposition of strategies). Beyond that, the notation and the details of formal development are completely in the hand of the real-time performing ecosystem.

There is an internal hierarchy, with the iPhone player on top (considering the privileged relationship with the score). To some extent, an AR drawing may be assimilated to a formal section (or to the whole piece) in a commonly notated composition: the “main idea” is the entire virtual painting, and the development lies in the different perspectives one can obtain (zoom-in, zoom-out, rotation, exclusion from the field of view, etc.).

3.9 Relation of the score with the audience and with the environment.

The essential feature of AR consists in blending the real environment with digitally rendered objects. The presence of an object in the real space clarifies its spatial existence and dimension (this is especially important if we recall the idea of virtual tangible scores presented in 3.5). The interaction with the real environment brings the score itself inside the space of the performance. The score has a 3-D inclusive nature (it can potentially include the entire venue of the performance). The audience is, in some way, part of the whole process of creation and can be surrounded by those virtual objects.

Figure 11. A performance (Liverpool, FACT 3).

Essentially, the projected score also acts as a visual part of a multimedia performance. While the VR mode presents only the score itself, AR also provides a perspective on venue and spectators. For this reason, the AR mode is considered the main one.

4. ISSUES AND LIMITATIONS

In its current form, LINEAR shows some limitations, ranging from the still preliminary stage of artistic development to the imperfections in positional tracking.

Regarding the latter issue, as explained in 3.2, ARKit makes use of feature points for performing its VIO algorithm. The absence of visual cues in the image detected by the camera will result in poor positional tracking. In worst cases, virtual objects move randomly around the scene.

Main causes for poor tracking are recognized as being bad light conditions and reduced visual complexity in the scene (therefore, lack of feature points) [20] .

Fast movements and sudden changes in the camera view easily lead to tracking issues (feature points must be compared between consecutive frames), reducing the iPhone performer's freedom of movement; at the current stage of development, the use of AR trajectories as choreographic indications, though promising, is not completely viable and presents risks for the stability of positional tracking.

Another problem is distance estimation. As [28] shows, users tend to underestimate distances, with obvious limitations to the flexibility and precision of interactions with virtual entities. Among many technical solutions, only shadows projected on the floor have a positive impact on distance estimation. At the current stage, downcast shadows are not implemented in LINEAR.

Even if the camera input is rendered on the screen, the device does not understand how the surrounding space is shaped, i.e., it does not understand depth data in the image. Therefore, the interaction of virtual and real world is still limited. For instance, virtual entities positioned behind a real object would not be hidden, as it would happen in reality (phenomenon called occlusion).

The possible notational solutions are currently constrained to only four body categories, each one emitting a particle effect. Even if an infinite number of different trajectories can be created, the look of a single virtual body or of a body category cannot change over time. The use of a VR mode makes the notational process more dynamical but does not overcome all the limitations.

There is also an intrinsic (and wanted) constraint: the system is not meant to create fixed, pre-composed AR scores. The author is currently working on another project aimed at filling this gap.

5. CONCLUSIONS AND FUTURE WORK

The current state of AR technology permits the exploration of unprecedented possibilities in musical notation and performance. While the technology itself existed for 50 years, only recently it has reached a level of flexibility and precision allowing a relative ease of implementation.

In this study, the author has presented a possible use of AR in LINEAR, where the OSC connection between an iPhone app, a Max/MSP patch and a streaming box produces an environment usable for performances based on live-generated animated scores and virtual interfaces. Its use sheds light on some concepts that have not been fully explored yet:

- virtual tangible scores (the iPhone performer plays virtual trajectories, i.e. the notation itself);
- notational feedback (some virtual bodies are created according to the analysis of the acoustic instruments' sounds; i.e., the notation is created by itself, as an effect of its reading);
- compositional ecosystem (all the performers have a direct influence on the notation and how it is interpreted).

In future works, notational process and performance strategies can undergo considerable improvements, especially with the design of a more complete and complex set of possibilities. Enhancements would include techniques of image processing for the camera input, such as distortion, frame differencing, tessellation as well as dynamic change of the visual features of virtual body categories throughout one performance.

The integration with different sensors could further expand the application functionalities. Major improvements in world tracking could be accomplished using 3-D ambient scanning sensors (as the *Structure Sensor* [21]). This implementation would allow a higher freedom in movement for the iPhone player (better positional tracking) and a better quality in the interaction between virtual and physical world. Introducing the use of a headset for mobile devices (*Bridge* [22] by *Occipital* for Apple devices) could also bring to a higher level of immersion for the iPhone performer and to a different approach with notation and perspective changes.

The idea of virtual tangible scores suggests the use of haptic devices in order to give the feeling of touch with virtual structures.

The continuous contact with performers (aimed at identifying limitations and at improving artistic and technical aspects of the system) is, and will be, an essential part of this research.

Acknowledgments

My PhD studies at Baptist University are funded by the HKBU UGC grant.

6. REFERENCES

[1] ARMusic, "Augmented Reality During Live Music Performances." [Online]. Available: https://www.youtube.com/watch?v=nyVs_5TfN4c

[2] L. Vickery, "Rhizomatic approaches to screen-based music notation," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, Brisbane, Australia, 2016, pp. 394–400.

[3] ——, "The evolution of notational innovations from the mobile score to the screen score," *Organised Sound*, vol. 17, no. 2, pp. 128–136, 2012.

[4] J. Freeman, "Extreme Sight-Reading, Mediated Expression, and Audience Participation: Real-Time Music Notation in Live Performance," *Computer music journal*, vol. 32, no. 3, pp. 25–41, 2008.

[20] Apple – *Introducing ARKit: Augmented Reality for iOS – WWDC 2017*: https://developer.apple.com/videos/play/wwdc2017/602

[21] Occipital, *Structure Sensor - 3D scanning, augmented reality, and more for mobile devices*: https://structure.io

[22] *Bridge - Mixed Reality and Positional Tracking VR Headset for iPhone and iOS*: https://bridge.occipital.com/

[5] G. Winkler, "The realtime-score. A issing-link in computer-music performance," in *Proceedings of the Sound and Music Computing conference (SMC'04)*, Paris, France, 2004.

[6] C. Hope, A. Wyatt, and L. Vickery, "The Decibel ScorePlayer: New Developments and Improved Functionality," in *Proceedings of the International Computer Music Conference (ICMC)*, Denton, USA, 2015.

[7] C. Hope, L. Vickery, A. Wyatt, and S. James, "The Decibel Scoreplayer - A Digital Tool for Reading Graphic Notation," in *Proceedings of the International conference on Technologies for Music Notation and Representation (TENOR'15)*, Paris, France, 2015.

[8] D. Kim-Boyle, "The 3-D score," in *Proceedings of the International conference on Technologies for Music Notation and Representation (TENOR'17)*, A Coruña, Spain, 2017.

[9] J. J. Ham, "An architectural approach to 3D spatial drum notation," in *Proceedings of the International conference on Technologies for Music Notation and Representation (TENOR'17)*, A Coruña, Spain, 2017.

[10] D. Hmeljak, "3D Musical Notation - Providing Multiple Visual Cues For Musical Analysis," in *Proceedings of the XII Colloquium on Musical Informatics*, Gorizia, Italy, 1998.

[11] B. E. Carey, "SpectraScore VR: Networkable virtual reality software tools for real-time composition and performance," in *International conference on New Interfaces for Musical Expression (NIME)*, Brisbane, Australia, 2016.

[12] H. Ling, "Augmented Reality in Reality," *IEEE MultiMedia*, vol. 24, no. 3, pp. 10–15, 2017.

[13] I. E. Sutherland, "A Head-Mounted, Three-Dimensional Display," in *AFIPS'68 Proceedings of the Fall Joint Computer Conference, Part 1*, 1968, pp. 757–764.

[14] F. Berthaut, C. Arslan, and L. Grisoni, "Revgest: Augmenting Gestural Musical Instruments with Revealed Virtual Objects," in *17th International Conference on New Interfaces for Musical Expression (NIME)*, Copenhagen, Denmark, 2017, pp. 180–185.

[15] I. Poupyrev, R. Berry, J. Kurumisawa, K. Nakao, M. Billinghurst, C. Airola, H. Kato, T. Yonezawa, and L. Baldwin, "Augmented Groove: Collaborative Jamming in Augmented Reality," in *ACM SIGGRAPH 2000 Conference Abstracts and Applications*, New Orleans, USA, 2000.

[16] I. Poupyrev., R. Berry, M. Billinghurst, H. Kato, K. Nakao, L. Baldwin, and J. Kurumisawa, "Augmented reality interface for electronic music performance," in *Proceedings of the International Conference on Human-Computer Interaction (HCI)*, New Orleans, USA, 2001, pp. 805–808.

[17] R. Berry, M. Makino, N. Hikawa, and M. Suzuki, "The Augmented Composer Project: The Music Table," in *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR'03)*, Tokyo, Japan, 2003, pp. 338–339.

[18] F. Farbiz, Y. T. Ka, K. Wang, W. Ahmad, C. Manders, J. H. Chong, and K. T. Yeow, "A multimodal augmented reality DJ music system," in *International Conference on Information, Communications and Signal Processing*, Singapore, 2007.

[19] T. Laurenzo, "YARMI - An Augmented Reality Musical Instrument," in *International conference on New Interfaces for Musical Expression (NIME)*, Pittsburgh, PA, USA, 2009, pp. 268–269.

[20] F. Liarokapis, "Augmented Reality Scenarios for Guitar Learning," in *EG UK Theory and Practice of Computer Graphics*, L. M. Lever and M. McDerby, Eds. The Eurographics Association, 2005.

[21] C. Kerdvibulvech and H. Saito, "Vision-Based Guitarist Fingering Tracking Using a Bayesian Classifier and Particle Filters," in *Advances in Image and Video Technology – Second Pacific Rim Symposium (PSIVT)*, Santiago, Chile, 2007, pp. 625–638.

[22] V. Chouvatut and W. Jindaluang, "Virtual piano with real-time interaction using automatic marker detection," in *International Computer Science and Engineering Conference (ICSEC)*, Nakorn Pathom, Thailand, 2013, pp. 222–226.

[23] E. Strasnick, A. Chambers, L. Jiang, and T. Xiaonan, "Pianolens: An Augmented Reality Interface For Piano Instruction," 2017. [Online]. Available: https://cs.stanford.edu/people/estrasni/research/pianolens/

[24] I. Bukvic and S. Lee, "Glasstra: Exploring the Use of an Inconspicuous Head Mounted Display in a Live Technology-Mediated Music Performance," in *International conference on New Interfaces for Musical Expression (NIME)*, Copenhagen, Denmark, 2017, pp. 313–318.

[25] E. Tomás and M. Kaltenbrunner, "Tangible Scores: Shaping the Inherent Instrument Score," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, London, UK, 2014, pp. 609–614.

[26] C. M. Fischer, "Understanding Animated Notation," in *International conference on Technologies for Music Notation and Representation (TENOR'15)*, Paris, France, 2015.

[27] I. Xenakis, *Formalized Music. Thought and Mathematics in Composition*. Pendragon Press, 1992.

[28] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir, "Designing for Depth Perceptions in Augmented Reality," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Nantes, France, 2017, pp. 111–122.