ACTION SCORES AND GESTURE-BASED NOTATION IN AUGMENTED REALITY

Giovanni Santini
Hong Kong Baptist University
info@giovannisantini.com

ABSTRACT

Augmented Reality (AR) is becoming, year by year, an established and well-known technological resource. Experimentations and innovative applications are produced in different areas. In music, there already is some use of such a technology in the fields of education and performance. However, the use of AR features as compositional resources has yet to be deeply explored and leaves room for innovative research. In particular, the possibility of notating the gesture in space, instead of on paper or screen, has been only superficially studied. This research focuses on the development of a new prescriptive notation system for gestures that represents extended techniques requiring direct contact between the performer and the vibrating body. Such a system has been implemented in the composition *Portale*, for small tam-tam, AR environment and live-electronics.

1. INTRODUCTION AND BACKGROUND

The notation of gesture for music scores has been at the center of numerous experimentations at least in the past 60 years. The need to expand the Common Music Notation (CMN) system derived from an enlarged aesthetic panorama which, in many circumstances, was taking into account sounds and performing techniques that had not been considered and included in the standard practices of music. The notation of gesture is often connected to the notion of prescriptive notation (indication of the action) as opposed to descriptive notation (description of the result).

1.1 Action scores

In Helmut Lachenmann's action scores, "…the score is notated as a series of actions without determining their precise pitch content or even their precise sounding result" [1]. In other words, it implies a notation of gesture which leaves room for some unpredictability in the sonic outcome. The panorama is extremely vast and varied and it is impossible to provide a satisfying background in this paper. Three examples from three different authors will be provided.

Copyright: ©2020 Giovanni Santini. 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.

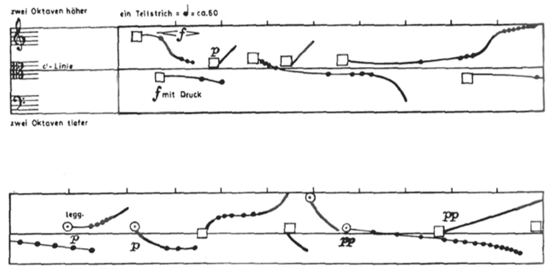

Figure 1: First two lines of Lachenmann's *Guero*

Figure 1 shows the beginning of H. Lachenmann's *Guero* (1969), where the pianist plays the keys themselves (as the emitter of the sound), with no action on the strings. The clefs report the whole keyboard extension (almost as a simplified geography of the instrument) and the score is based on gestural information (e.g., white squares correspond to the white keys played with finger's nails) and relatively free proportional time indications. Some degree of unpredictability is intrinsic in the notational system itself and is regarded as an aesthetic value of the compositional thinking.

A further evolution of gesture-based notation, almost completely excluding CMN elements, can be found, for example, in compositions by Aaron Cassidy, such as the *Second String Quartet* (2010), where notes are completely replaced by lines indicating the left hand's fingers' positions and indications such as trill and vibrato. Moreover, a grey scale is used to deliver pressure information (from full pressure to harmonics pressure). A red line indicates bow position, pressure (using transparency) and strings of contact. A green line indicates the bowing (which portion of the bow is being pressed on the string at a given time). Rhythm of left and right movements is notated on additional staves. The result is a complex and multilayered tablature that "examines the ways in which limited collections of physical action types can 'push against' constructed, dynamic, multi-planar bounding windows". The complexity of their combination "encourages unusual, unexpected, and often unpredictable materials to emerge" [2].

Pierluigi Billone's *Mani.Mono* (2007) (Figure 2) includes drawings of the gestures to perform along with the symbolic prescriptive notation of the gesture and the corresponding descriptive notation of the intended sound result. In this case, gesture-based notation is oriented towards the sound in a very specific and "deterministic" way: there is one expected sound result which is meant to be consistent across different performances. The score indicates how to

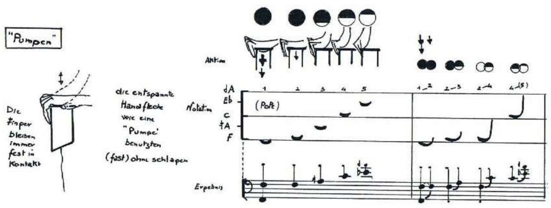

Figure 2: The beginning of Billone's *Mani.Mono*

reach a specific result rather than constructing a set of possibilities through a parametric representation of gesture.

In all the mentioned examples, the notation is coherent with the aesthetic purposes of the composers and is adequate to meet the artistic goals. This research is not meant to point out limits of 2D notation in general. However, new perspectives beyond the already known uses of notation can be pointed out: notation on paper would always show limits in carefully indicating spatial information, as it requires some form of abstraction from the space itself. In other words, notation on paper can't notate spatial events exactly where they are supposed to happen, with their
4-dimensional behaviour. Conversely, AR notation has this capability. As discussed in sections 2 and 4, in cases where an articulated use of gesture needs to be connected with a precise and consistent sound result (with the intention to minimize the intrinsic unpredictability of prescriptive notation), AR can deliver an unprecedented level of precision.

1.2 AR for music education and performance

Before describing the content of the AR-based notation system developed by the author, it is worth pointing out some background research in Augmented Reality for music education, composition and/or performance.

In music instruction, differentiated areas of interest have been taken into consideration. Interest have arisen around Western instruments, such as guitar [3, 4, 5], violin [6, 7], and non-Western instruments such as Guqin [8], Koto [9], and Dombyra (Kazhakh traditional instrument) [10]. The most researched instrument is piano, addressed in a large number of papers (e.g., [11, 12, 13] to cite a few). The piano roll is a quite standard tool: virtual blocks appear in correspondence of the keys to press. As long as a block is visible, the corresponding key has to be held down. When the block disappears, the key has to be released.

According to evaluation studies carried in the cited research papers, AR lowers the barrier of entry for beginners, guarantees higher accuracy and better mnemonic retention. The target audience of applications are students in very early stages and the proposed repertoire includes compositions studied in the first one or two years of learning. Although benefits for beginners have been proven, experts tend to find AR notation systems on traditional repertoire confusing and impractical.

It is important to notice that all the mentioned applications have been developed for allowing students to learn an already existing repertoire, as an aid to traditional instructional means more than as a stand-alone solution.

With the release of devices dedicated to AR, some experiments have risen. In [14], Kim-Boyle presents the concept of immersive score. The score *5x3x3*, previously realized in 3D, is ported into an AR environment and in room-scale size. The performer, wearing the *HoloLens 1* [1] headset, is immersed in a virtual structure (the score), superimposed to the real world and can navigate it. In this context, AR is used mainly in its visual capacity; however, different properties of the score react to timbral nuances, thus introducing interactive functions in the AR notation.

On the other hand, the works of Amy Brandon, such as *Augmented Percussion* [15], present a deeper focus on AR interfaces and interaction: bare hands of a percussion performer are used to activate/manipulate the sound of virtual objects, some of which are embedded in the real instrument. The interpreter makes use of a *Meta 2* [2] headset.

LINEAR [16] is an AR framework for improvisation that allows the creation of a notation-interface hybrid in real-time: a performer, using an iPhone, can generate virtual bodies along the trajectory of his/her gestures. Those bodies have both the function of interface (they are linked to specific samples) and notation (for generating the sound of the original gesture, the performer needs to repeat the same gesture).

[17] describes an AR controller for sound spatialization in n-channels designed both for diffusion automation and for live performance. It allows to position virtual bodies in the space and to draw trajectories with the gesture. Each of these bodies is linked to a sound source moving in the space along the designed trajectories. The tool is designed for HTC Vive Pro [3] .

2. THE AR ACTION SCORE: WHY AND HOW

The concept of AR action score will be explained by analyzing the particular case of the author's *Portale*(2019), for small tam-tam, AR environment and live-electronics.

In the composition, the tam-tam is played in 2 ways:

- by using fingers wearing thimbles;
- by moving one magnet (held in position by another magnet in the back) over the surface of the instrument.

The finger generates a scratchy timbre when it is moved continuously on the surface. It also produces a more "pitched" sound when hitting the tam. The magnets move the perceived pitch and shift spectral components, behaving as masses attached to the vibrating body. The will to realize a clear and intuitive notation for indicating the movement of the magnets (addressed to a specific spectral configuration) is at the core of the development process.

[1] An AR see-through device developed by Microsoft.
[2] An AR see-through device developed by MetaVision.
[3] Non-see-through head-mounted display headset allowing AR content thanks to the use of front-facing camera for visualizing the real environment

2.1 How the magnets modify the tam's spectrum

Figures 3a and 3b report two different results (spectrogram and pitches of the loudest partials) generated by the percussion of a finger (with thimble) hitting one specific tam-tam in a specific position (white cross) with the magnet(s) in another specific position (white circle).

Although some difference has to be expected among repeated hits, the spectral content of sounds is consistent for every position of the magnet and of the hitting finger, given a specific tam-tam, a specific magnet and a specific thimble. Figure 4 shows the spectral shift obtained with a magnet moved on the tam-tam. Every trajectory performed on the surface of the instrument results in a different shifting effect (although with some similarities between similar movement shapes).

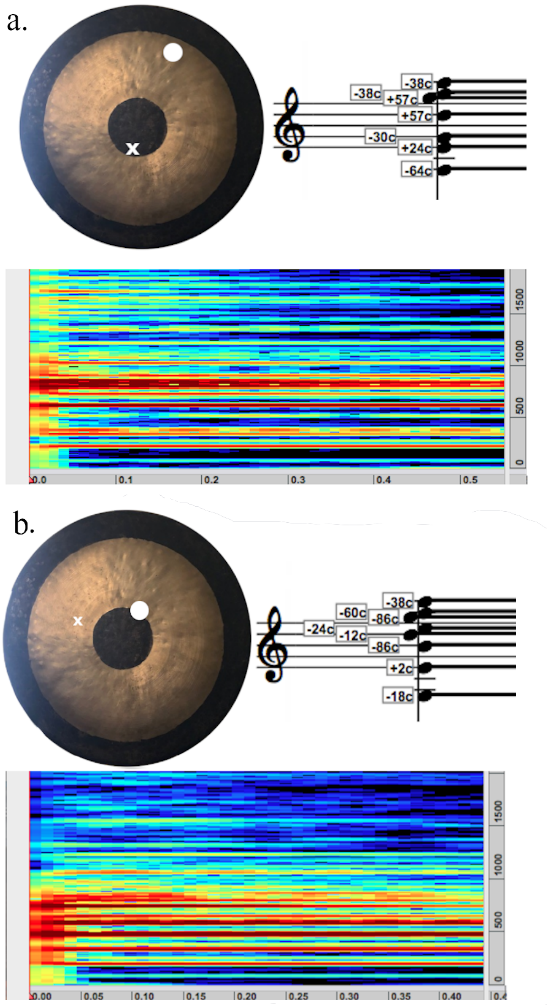

Figure 3: Analysis of the excitation of the tam by a metal thimble at the location indicated by the white X with the magnet at the location of the white circle. 3a and 3b indicate two different positions both for the thimble and for the magnet. Spectrogram shows the first 0.6s of the attack, from 0-2kHz. The seven highest amplitude frequencies are showed in common notation, with cent deviations.

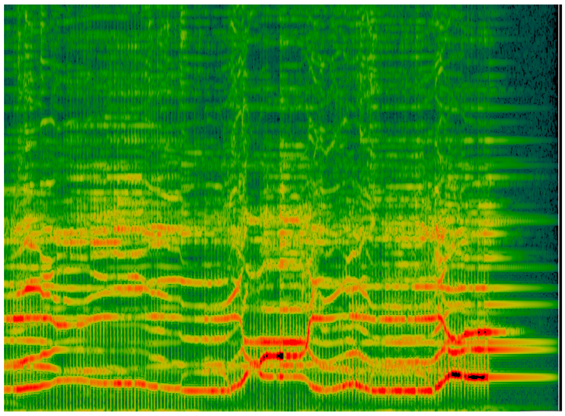

Figure 4: Spectrogram displaying the shift of partials during the movement of the magnet. When the magnets stop moving the partials become visibly more stable in pitch distribution.

2.2 Notate the effect by notating the action

In general, the accurate identification of sound results in the context of extended techniques poses well-known limitations [4]: instability, unpredictability and limited control over the result (although inside a well-defined timbral world).

The rationale behind the notation system presented here is that, by being able to indicate the specific point in space where the interaction has to be realized, it is possible to increase the level of accuracy, while decreasing the complexity of the deciphering process.

In fact, if any position of the magnet and of the thimble on the instrument correspond to a quite consistent class of spectra, the identification of the precise point of access is enough for also indicating the result with a satisfying precision. In addition, the notation is almost self-sufficient, requiring few preliminary information. On the contrary, a notation on paper could need to be complemented with additional audio material and performance notes (especially in circumstances where a precise spectral result is required).

The notation system implemented consists in virtual points indicating hit spots and in virtual lines following pre-designed trajectories projected directly on the tam-tam. The performer can see those objects by wearing a headset for AR rendering.

Figure 5 shows the two playing modes on tam (with fingers and with magnets). The line in the picture on top illustrates the trajectory the performer has to follow on the tam with the index finger. Such a line is not static but moves (therefore, the direction to follow is made obvious by the direction of the line along the trajectory). The picture at the bottom presents the AR indication corresponding to the configuration presented in Figure 3a: the AR model of the magnet (black virtual sphere) indicates the intended magnet position and the light blue effect representing the hit point for the finger.

[4] In some cases, considered as aesthetic resources.

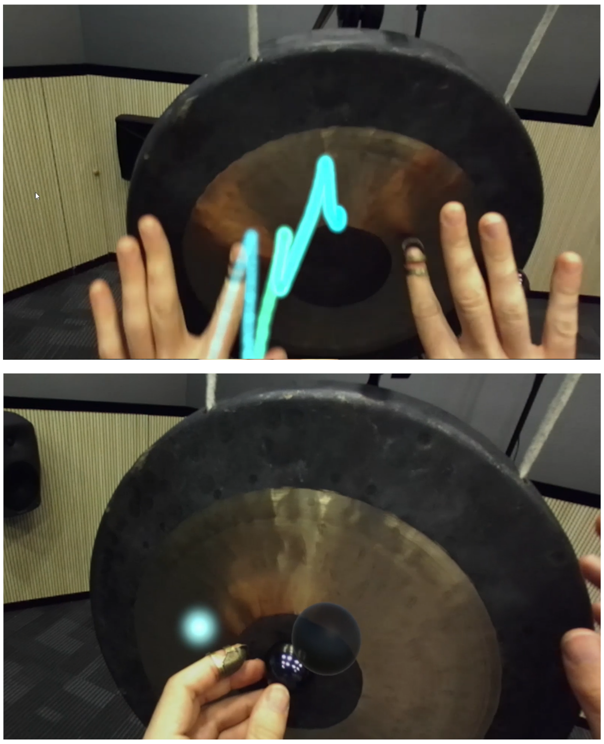

Figure 5: The two ways the performer is meant to play the tam: fingers with thimbles and magnet(s). On top, the AR line representing a trajectory the performer has to follow. At the bottom, the AR notation of the configuration in Figure 3a.

2.3 Notation of time and notation of intensity

In a prescriptive notation system as above, that consists in a line mimicking the gesture to perform, the notation of rhythm has two specific aspects.

The first one is that the notation is not symbolic (it does not use a figure that resembles something else from itself) but mimic (the movement is actually performed in advance by a virtual object and subsequently repeated by the performer). We are used to adopt some kind of symbol (or position on paper) to resemble durations (or time proportions). In general, we divide greater time values in smaller ones, or create longer durations by adding up shorter ones. In the AR notation system introduced here, time is not represented as multiples (or fractions) of a fixed amount of duration; on the contrary, it is represented as a fluid alternance of internal speed articulations of the gesture. We could call it *continuous rhythm* as opposed to *discrete rhythm* (the rhythm expressed by metrical values). Such fluid notation of time can take into account the differences of speed inside a gesture: in fact, a real performance act can have an unstable velocity, connected to differences in sound result (e.g., different speeds of the bow on a string or, as in Portale, different speeds of a magnet on a metal surface). Although the metric notation in CMN can become extremely specific in defining different durations, it will always describe the articulation of a gesture as a sequence of finite durations (for going from a point A to a point B) and not as a fluid change in the gesture's velocity.

The second aspect is related to the performer's reaction time. In fact, rehearsed rhythm learned on a score is precise. On the contrary, in this system the information is conveyed to the interpreter in real-time and every action is notated in the moment in which it should happen. Clearly, the performer cannot realize the required actions as soon as they appear on the instrument. The execution needs some delay time. For this reason, a certain fuzziness in the rhythmical outcome is an intrinsic component of such a notation system.

At the current stage there is no clear indication of dynamics (which has a noticeable impact on the spectral result); adding some indicators for that parameter is a forecasted enhancement of the software. However, the system already notates width and speed of movements and hits, which have a close relation with dynamics (faster and wider movement for *ff*, opposed to small and slow movements for *pp*).

3. TECHNICAL ASPECTS

The notation system used in *Portale* relies on a technical environment requiring different hardware and software components. In addition to microphones, speakers, sound interfaces and pc(s) for audio, visual and positional tracking processing, the framework requires:

- 1 Head Mounted Display (HMD) [5] ;
- 1 stereo VR camera [6] ;
- 2 motion capture trackers [7] ;
- a software developed in Unity 3D for AR processing.

3.1 Headset and trackers

The performer is wearing the HTC headset, which allows the representation of virtual bodies in space by detecting the real world with front-facing cameras and representing it on the internal screens (one per eye). Virtual bodies are rendered on the same screens. The front-facing cameras natively installed on the headset deliver a poor image quality (420p per eye) and have a high latency (200 ms), making it problematic for the performer to follow the notation accurately. For this reason, the *ZED Mini* Stereo VR Camera has been mounted on the headset for replacing the native one, bettering the resolution to 720p per eye and lowering the latency to 60 ms.

The two trackers are positioned:

- on the tam for detecting the position and orientation of the instrument. This way, the AR score always follows the tam's movement;
- on the right hand of the performer. This tracker is used as an input device for interacting with virtual bodies (this component of the composition is not a part of the notation system described in this paper and therefore its function will not be further addressed).

[5] *HTC Vive Pro* headset.
[6] *ZED Mini* VR Camera.
[7] *Vive Trackers*, devices used for detecting the position and orientation in space of objects in the real world.

3.2 Software

The software side of *Portale* is articulated in two components:

- an AR program created and compiled in Unity;
- a Max/MSP project (its functioning is not further analyzed as it is not implied in the notation system).

The AR software is responsible for scheduling, rendering and positional tracking processing. The software also sends control information to Max/MSP when the interaction of the physical performer with virtual objects (detected through motion tracking) is meant to produce some sound outcome (generated in Max/MSP).

3.2.1 Gesture design

In the application, gestures are resembled by a virtual object (the blue line in Figure 5) following a trajectory with a certain speed.

The creation of an Augmented Reality score poses two issues: how to draw a 3D gesture and how to move a body along that trajectory in time.

The trajectory of the gesture is created with a 3D cubic Bezier's curve (a parametric curve used in computer graphics), whose shape and bending can be adjusted by shifting the position in space of nodes (the white squares in Figure 6) and control points (the red squares in Figure 6, two per node). Trajectories are designed in advance and cannot be changed during the performance.

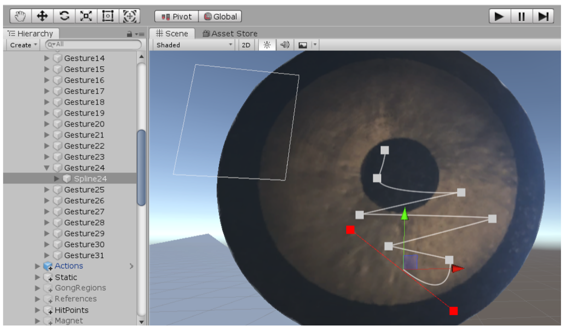

Figure 6: The creation of one trajectory with a sequence of cubic Bezier's curves inside the Unity editor.

Two *scripts* [8] allow the composer to control the starting point in time, the internal speed articulation of the gesture (how long does it take for the line to go from one node to the next) and its total duration.

4. DISCUSSION

The concept of AR action score and AR gesture-based notation is conceived in the frame of extended techniques and timbral research. Its main focus lies in the delivery of mimic (not symbolic) prescriptive information in 3D space which, while ensuring a certain degree of intuitiveness, allows an accurate control over the result. In fact, in every circumstance where the result is consistent given a specific position of exciters and/or preparation [9], the system provides a univocal way of notating that result (as a function of the positions or gestural behavior). The particular nature of time indication in this context produces a condition which could be defined as continuous rhythm, as a consequence of its capability of notating speed changes in gesture instead of events happening in relation to fixed rhythmical values or in relation to time proportions linked to the position on paper.

In *Portale*, the formal development is focused on the evolution of different forms of interaction between the physical performer and a virtual object (the blue animated line that runs across trajectories) which, in the last section, becomes AR gesture-based notation. Describing the other stages of interaction would go beyond the scope of this paper. However, it is worth mentioning that, in its implementation for *Portale*, the notation system has been used for a restricted set of pre-designed actions with a relatively narrow space for development: only two playing techniques, only one trajectory at a time, the impossibility for the physical performer to interact with the notation itself (but only follow it). These constraints have been implemented mainly for two reasons:

- *Portale* is structured around different possibilities of interaction between the physical performer and virtual objects (and viceversa) and, in this context, AR notation is one of them;
- the physical conformation of the small tam did not allow to safely and/or effectively use some solutions [10].

These constraints are not to be considered limitations of the system itself, but rather choices of implementation. Other versions of the system for future compositions focusing exclusively on AR gesture-based notation will include broader sets of possibilities.

Although, in the opinion of the author, such a quite unprecedented possibility shows potentials for future musical research, there are some intrinsic limitations.

While the system can be considered precise in static situations, when movements (especially for the magnets) are considered, the instrument itself and the magnets attached to it oppose some resistance to the performance gestures. In fact, the irregularities of the surface of the tam might sometimes prevent the magnet from keeping its position or following the desired path. As a result, a precise indication of movement does not automatically translate into a precise movement.

Another issue consists in the difficulty of adaptability of the notation. For example, playing *Portale* on a tam different from the one on which the composition has been developed, would make the sound result (at least slightly) different. In fact, every tam, even if of the same size and

[8] A script is a custom programming file written in C# that can be attached to virtual objects in order to control their behavior.

[9] Preparation is here intended as a modification of the usual behavior and timbral result of a vibrating body by the addition of extraneous masses to the vibrating body itself.

[10] E.g., the use of two sets of magnets could have created problems given the small size of the instrument. In fact, there is a high chance of having the two sets near enough to generate magnetic attraction, thus making impossible to have the control over the required techniques

from the same constructor, could present significant timbral divergences. Those dissimilarities would be particularly obvious in case of tams with different sizes. This limitation would also hold for other instruments on which this notational system might be applied.

All of this said, the notation system itself would not present particular problems of scalability in size (e.g., changing from an 18" to a 32" tam): the same score would easily fit a bigger or smaller instrument. That is because the reference system can be scaled inside the AR software, and therefore the score itself and all the gestures can be instantly resized accordingly.

One additional constraint is constituted by the organological nature of the instruments on which the system could be used. It would only work for instruments whose vibrating part can be directly manipulated by the performer with the gesture (surface-like instruments): most of percussion instruments and strings; conversely, it could not fit for woodwinds and brasses (where the vibrating body is constituted by the air column).

An informal evaluation carried on December 1st 2018 at the *AR/VR Retreat* in Berkeley (CA) showed that non-musicians were able to perform a 30 seconds AR score producing a result reasonably close to the intended outcome after just one repetition after an instruction process that lasted around one minute. Although only a structured evaluation (which will be presented in future research) could provide trustable results, the informal one was encouraging.

5. CONCLUSIONS

The notation system presented in this article, developed for the composition *Portale*, allows one to draw in space and time the intended gesture addressed to a specific sound result. The score is formed by a series of static points (indicating specific spots on the instrument) and lines moving along pre-designed trajectories rendered on the tam. This notational system guarantees to preserve immediateness and intuitiveness of notation, while being accurate on the result. In fact, as shown in Session 2, it is possible to compose the harmony derived from the movement of the magnet across the surface of the tam without needing extra sound material or additional spectral/harmonic information. The indication of the position of interaction or of the trajectory is sufficiently accurate to determine quite consistently the result.

Another point of interest can be found in the notation of rhythm, not realized through symbols referring to discrete values (such as quaver), or through a proportional graphic distribution, but coming from the internal speed articulation of the gesture. AR notation, as implemented in *Portale*, mimics the behavior of gesture over time and represents visually and in real-time the fluid alternation of velocities with a level of similarity that notation on paper could not possibly achieve. Such a particular dimension of temporal indication might be referred to as *continuous rhythm*.

The system requires the use of a specific headset for AR rendering (HTC Vive) and is realized through a custom software created in Unity 3D. The trajectories used for the score can be generated with a graphic editor and custom scripts allow to compose the internal speed of movement on each trajectory. Trajectories are then automatically placed on the real tam through the use of the Vive Tracker.

Main limitations consist in difficulties of adaptability (different tams provide a slightly to greatly different sound result given the same interaction positions and trajectories). Additionally, the notation is fruitful only on instruments providing a surface for interaction (while it could not work on instruments using the air column).

In future work, the realization of a formal evaluation experiment will provide more information on the actual usability, precision and intuitiveness of the system.

6. REFERENCES

[1] A. Heathcote, "Liberating sounds: phylosophical perspectives on the music and writings of Helmut Lachenmann," Ph.D. dissertation, Durham University, 2003.

[2] A. Cassidy, "Constraint Schemata, Multi-axis Movement Modeling, and Unified, Multi-parametric Notation for Strings and Voices," *Journal for new music and culture*, 2013.

[3] C. Kerdvibulvech and H. Saito, "Vision-Based Guitarist Fingering Tracking Using a Bayesian Classifier and Particle Filters," in *PSIVT'07 Proceedings of the 2nd Pacific Rim conference on Advances in image and video technology*, vol. 14, 2007, pp. 625–638.

[4] M. Löchtefeld, S. Gehring, R. Jung, and A. Krüger, "guitAR: supporting guitar learning through mobile projection," in *CHI 2011*, 2011.

[5] J. R. Keebler, T. J. Wiltshire, D. C. Smith, and S. M. Fiore, "Picking up STEAM: Educational implications for teaching with an augmented reality guitar learning system," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.

[6] F. D. Sorbier, H. Shiino, and H. Saito, "Violin Pedagogy for Finger and Bow Placement using Augmented Reality," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2012.

[7] H. Shiino, F. De Sorbier, and H. Saito, "Towards an augmented reality system for violin learning support," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.

[8] Y. Zhang, S. Liu, L. Tao, C. Yu, Y. Shi, and Y. Xu, "ChinAR: Facilitating Chinese Guqin learning through interactive projected augmentation," in *ACM International Conference Proceeding Series*, 2015.

[9] M. Doi and H. Miyashita, “Koto learning support method considering articulations,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.

[10] M. K. Zhaparov and U. Assanov, “Augmented reality based on Kazakh instrument ”Dombyra”,” in *8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014 - Conference Proceedings*, 2014.

[11] M. Weing, A. Röhlig, K. Rogers, J. Gugenheimer, F. Schaub, B. Könings, E. Rukzio, and M. Weber, “P.I.A.N.O.: Enhancing Instrument Learning via Interactive Projected Augmentation,” in *2013 ACM Conference on Ubiquitous Computing, UbiComp 2013*, 2013, pp. 75–78.

[12] X. Xiao, B. Tome, and H. Ishii, “Andante: Walking Figures on the Piano Keyboard to Visualize Musical Motion,” in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2014)*, 2014, pp. 629–932.

[13] E. Strasnick, A. Chambers, L. Jiang, and T. Xiaonan, “Pianolens: An Augmented Reality Piano Interface.” [Online]. Available: https://www.youtube.com/watch?v=5TExa2L1rOM

[14] D. Kim-Boyle and B. Carey, “3D scores on the HoloLens,” in *TENOR 2019 International Conference on Technologies for Musical Notation and Representation*, Melbourne, 2019.

[15] A. Brandon, “Augmented Percussion,” 2019.

[16] G. Santini, “LINEAR - Live-generated Interface and Notation Environment in Augmented Reality,” in *TENOR 2018 International Conference on Technologies for Musical Notation and Representation*, Montréal, 2018, pp. 33–42.

[17] ——, “Composing space in the space: an Augmented and Virtual Reality sound spatialization system,” in *Sound and Music Computing 2019*, Malaga, 2019, pp. 229–233.