Augmented Piano in Augmented Reality

Giovanni Santini
Hong Kong Baptist University
Ho Sin Hang Campus
Kowloon Tong, Hong Kong
info@giovannisantini.com

ABSTRACT
Augmented instruments have been a widely explored research topic since the late 80s. The possibility to use sensors for providing an input for sound processing/synthesis units let composers and sound artist open up new ways for experimentation. Augmented Reality, by rendering virtual objects in the real world and by making those objects interactive (via some sensor-generated input), provides a new frame for this research field. In fact, the 3D visual feedback, delivering a precise indication of the spatial configuration/function of each virtual interface, can make the instrumental augmentation process more intuitive for the interpreter and more resourceful for a composer/creator: interfaces can change their behavior over time, can be reshaped, activated or deactivated. Each of these modifications can be made obvious to the performer by using strategies of visual feedback. In addition, it is possible to accurately sample space and to map it with differentiated functions. Augmenting interfaces can also be considered a visual expressive tool for the audience and designed accordingly: the performer's point of view (or another point of view provided by an external camera) can be mirrored to a projector. This article will show some example of different designs of AR piano augmentation from the composition *Studi sulla realtà nuova*.

Author Keywords
NIME, Augmented Reality, Augmented Instruments, Music Performance

CCS Concepts
•**Applied computing → Arts and humanities**; Performing arts • **Applied computing → Arts and humanities**; Sound and music computing •**Human-centered computing → Human computer interaction (HCI) → Interaction techniques**; Gestural input

INTRODUCTION AND BACKGROUND
From the end of the 80s, researchers started to explore the possibilities of letting performers of acoustic instruments control electronics (Augmented Instruments) or translate the virtuosity of playing techniques in the digital domain (Hyperinstruments).

More in detail, innovative "approaches are possible with 'augmented instruments'. First, sensors can be utilized to add control possibilities that are not directly related to normal playing techniques. For example, various buttons can be added to the body of the instruments[…]. Second, sensors can be applied to capture normal playing gestures"[1]. In practice, augmented instruments increase the control a performer can have in performances implying electronics, by detecting his actions through sensors or physical interfaces.

"The basic concept of a hyperinstrument is to take musical performance data in some form, to process it through a series of computer programs, and to generate a musical result. […] The performer uses more-or-less traditional musical gestures, on a more-or-less traditional musical instrument. The instrument itself, however, is "virtual", since the computer system supporting it can redefine meaning and functionalities at any point"[2]. A similar concept can also be found in the definition of Digital Instrument (for a detailed definition, see [3]).

Research on Augmented Instruments and Hyperinstruments is essentially based on the idea of integrating electronic systems inside the musical performance with the aim to overcome the distinction between electronic processing (computer-generated music) and acoustic/gestural/human playing. The use of sensors or of newly designed interfaces for sound synthesis and parametric control of signal processing were developed in order to deliver a real-time (and usually intuitive) control over electronic sound production.

The newest possibilities offered by the means of VR and AR can further expand this kind of research, while providing a new, immersive framework.

A good number of applications have already been developed in the domain of digital VR Instruments [4]. For example, *ChromaChord* [5] allows the user to play virtual interfaces by using *LeapMotion* and *Oculus Rift*: colored blocked generated pitches when the hands of the user collide with those virtual objects. Another example is the *Synesthesizer* [6], a VR color-based synthesizer that uses Machine Learning to map colors scanned in a VR environment to a large variety of timbres. In *AirPiano*[7], the simulation of the real functioning and haptic feedback of a real piano keyboard is simulated inside a virtual environment. *VRMin* [8] is a VR augmentation for the Theremin. The principle behind it is similar to the one that is at the basis of the AR interface design presented here: the virtual representation in space of the interactive areas is used in order to increase the playing precision. Contrarily to the AR design for interfaces described in this paper, VRMin is in VR, i.e. the real instrument is not visible. Additionally, VRMin could be considered more as a learning tool than a performance tool.

Music education has also been considered for interactive AR applications featuring interfaces. In particular, the piano has attracted numerous researchers, e.g. in [9]–[12]. A quite recurrent strategy is the *piano roll* (or VR/AR MIDI roll), consisting in moving virtual blocks indicating the keys to press with the right timing.

Numerous applications have also been developed for live performance. LINEAR[13] is a tool for generating interfaces-notation hybrids in real-time (the trajectory of the gesture to perform is indicated by virtual bodies that produce the intended resulting sounds when "hit" by the performer). That is one of the first examples of accurate space sampling, where the exact position of interaction can be easily determined and used as a

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

technical and expressive resource. In [14] a Virtual Instrument for sound spatialization is described. The tool allows one to draw trajectories in the air (with embedded speed information) and place sound sources on those trajectories. The position over time of those sound sources is then simulated by a n-channel audio system. A. Brandon's *Augmented Percussion* (2019) is a composition where a marimba is augmented by placing virtual objects around it. Some of them are also embedded in the body of the instrument. Those objects are used to manipulate the processing of the marimba's sound. G. Santini's *Portale* is a composition for tam-tam, AR environment and live-electronics where virtual interfaces, AR augmentation and AR gesture-based notation are featured. Furthermore, the composition is articulated around different interaction possibilities between the real performer and a protagonist Virtual Object. Innovative solutions have been also explored for the audience, conceiving AR not only as a tool for the user/performer but also as a visually expressive component for an audience. Among the most interesting solutions, remarkable is the use of holographic film for giving a hologram-like perception without the use of headsets for the audience [15], [16]. The creation of innovative performance formats created for an AR fruition by people on their mobile devices has also been experimented: *AR-ia* [17] is a system for generating 3D live figures for the live rendering of an AR opera on high-end mobile phones.

All the mentioned explorations show a progressively developing panorama, where experiments contribute, year by year, to the development of best practices. This paper contributes by describing different design solutions for AR piano augmentation, a topic that still needs to be explored in depth.

1. ENVIRONMENT

The composition *Studi sulla realtà nuova* has been developed for a combination of hardware and software components. Beyond the traditional audio equipment (such as microphones, speakers and pc), the environment requires:

- One non-see-through Head Mounted Display (HMD)[1];
- One stereo VR front-facing camera[2];
- 3 motion capture trackers (one for detecting the position of the piano, the other two for the hands of the performer)[3];
- Software developed in Unity 3D for AR rendering and interaction; through Open Sound Control (OSC), sends commands to:
- Ableton Live making use of custom Max Effects (created through Max for Live) and software synthesizers and samplers.

The point of view of the performer can be mirrored to a projector, in order to share the AR view with the audience.

2. DIFFERENT TYPES OF INTERFACES

The interfaces used in *Studi sulla realtà nuova* can be divided into 4 different categories.

2.1 Object-trigger

The simplest form of AR interface is constituted by isolated virtual objects that can be used for triggering one specific action (a processing preset, a sample etc.).

Figure 1. A virtual interface (green sphere) used for triggering samples.

The interface in Figures 1-2 consist in a sphere that can be activated with a gaze interaction: a smaller sphere, followed by a tail particle effect (with blue and light blue particles in figure 1-2), indicates the center of the point of view of the performer, facilitating the right orientation of the performer's head. A *raycaster*[4] is used for determining whether an interaction with the interface occurred (if the *raycaster* intersects its surface, the virtual object is activated). A sort of explosion is used as a visual feedback for the performer and as an expressive idea for the audience, who can relate the visual effect to a modification in the musical content (in this case, a chord is changed in the software synthesizer).

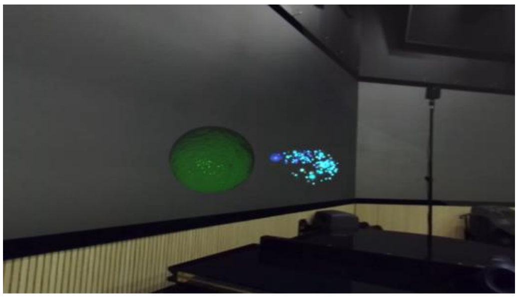

Figure 2. Visual feedback of the gaze interaction.

In another case, a water-like surface, "submerging" the keyboard, is used for triggering the next spectral delay preset in a Max for Live effect.

The object is activated when a hand emerges from the water (the hands' position is detected through the trackers). The visual feedback is delivered through the creation of reflections simulating some water's undulatory motion.

Figure 3. Water-like surface/interface

The dimension of the interface is disproportioned to its limited effect: a one-action trigger could have fitted a much smaller space. However, in this case, the design has privileged the scenic

[1] *HTC Vive Pro* headset.
[2] *ZED Mini,* providing a better quality and a shorter latency than the headset's original front-facing VR cameras.
[3] *Vive Trackers*.

[4] A test of intersection with surfaces. In this case, a straight line sent from the center of the point of view towards infinite in the direction of the gaze.

and visually expressive quality over the functionality (the interface is designed for the audience's point of view).

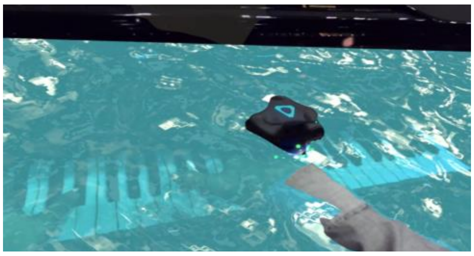

Figure 4. Visual feedback of the interaction

2.2 Multi-dimensional sliders

Another way to design a virtual interface is to map coordinates in one, two, or three dimensions in space to a continuum/multiple *continua* of values used for controlling parameters, parameters' groups and interpolated presets. Rotation values can be used for gaining an additional set of three dimensions (x, y and z axis values).

In *Studi sulla realtà nuova*, a tracker is used to access the cartesian coordinates of the hand's position[5] in space, the Euler angles of its rotation over three axis and the rotation Quaternion (a vector constituted by one real and three complex numbers used for encoding rotations in 3D space).

Therefore, it is possible to use data gathered from the tracker sensor for controlling multi-dimensional sliders. Those kinds of virtual interfaces are activated when the performer's hands enter inside the portion of space occupied by the virtual body. Upon activation, positional data are used to control the audio software. When the performers' hands are outside of the interface, positional data are not sent.

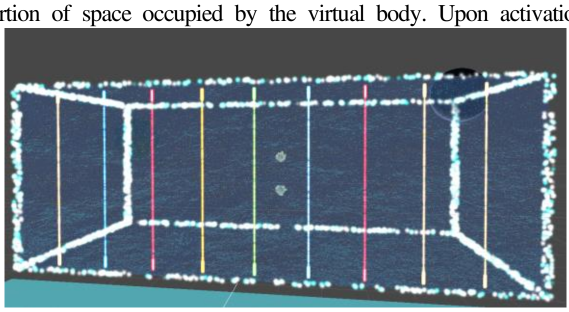

Figure 5. A multidimensional slider, shaped as a block of ice, with frets for different notes.

For the interface in Figure 5, the x-axis (left-right) is used for pitch (quantized to 12-tone equal temperament, indicated by the frets), y-axis (down-up) for microtonal tuning and z-axis for loudness (rear-front). The rotation of the hand around the z-axis is used for controlling a low pass filtering cutoff frequency.

2.3 Composite interfaces

A more articulated way for augmenting the piano in *Studi sulla realtà nuova* consists in creating a composite interface constituted by a series of interactive points in space (triggers) and/or interactive portions of space (sliders). Those single objects are then arranged around the instrument. By delivering a larger set of input points, a composite interface lets the interpreter completely control the electronics processing, without the need of any external help (e.g. for triggering live-electronics cues).

The use of this kind of interface can also allow a new degree of virtuosity, as the different virtual bodies can be used and seen as additional parts of the instrument itself, exactly located in space, with which the composer can require an accurate and virtuosic interaction over time.

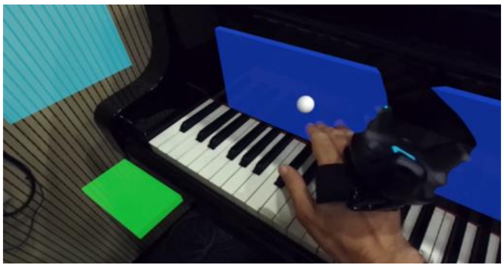

Figure 6. An example of composite AR inteface.

The interface in Figure 6 is composed by a series of triggers (green and light blue bodies, arranged around the instrument) and a 1D slider (the light blue body, positioned perpendicularly over the keyboard) modifying the intonation of a chorus effect.

2.4 Interface-notation hybrid

Multiple virtual bodies can also be positioned along a trajectory in space and their appearance can be consecutively delayed so that the required speed for following the trajectory is also suggested. In this case, a 4D (3D space + time) indication of the gesture to perform is carried out by virtual objects which, at the same time, can be assigned to some control actions. We can call interface-notation hybrid such a combination of gestural information and control interface.

In figure 7, the performer follows the trajectory indicated by the virtual bodies and hit the string at the end of such a trajectory. The speed of movement suggested by the consecutive appearance of the virtual spheres is also an indication of loudness of the final string hit. At the same time, when the hand passes through the position of the virtual bodies, it destroys those objects and triggers sound samples, except for the last body (the nearest to the piano strings) which triggers the new preset for a combination of chorus, granular synthesis and spectral delay. The disposition in space of the bodies indicates a gesture to perform. That gesture, once performed, leads, on one side, to hit the string with a certain speed (therefore loudness), on the other side, it produces sounds on its own because of the samples triggered by the virtual bodies. The movement, even before hitting the string, is, by itself, finalized to the production of

Figure 7. An interface-notation hybrid. The performer is starting following the trajectory which will end on piano strings in the middle register (meant to be hit with the hand).

[5] More precisely, the local position, i. e., the position respect to the central point of the interface.

sound. Therefore, the placement of virtual bodies notates the position in space both for virtual and for real sound-generating hits. Consequently, while being interfaces, the virtual bodies also constitute a form of (gesture-based) notation.

The interface-notation hybrid can be also generated by the performer, through his/her own gestures. In this case, virtual bodies are instantiated at regular time intervals along the trajectory created by the performer before hitting the string.

3. THE SCORE

Beyond interface-notation hybrids, *Studi sulla realtà nuova* requires the use of traditional notation. In those cases, a 2D score needs to be used. As the performer is wearing a headset, such a need poses some problems.

In fact, the HTC Vive Pro delivers visuals rendered on screens (one per eye) of the real world through the front-facing cameras. Consequently, the resolution of camera and screens is crucial. The camera provides a 720p per-eye resolution, which makes reading real scores quite problematic (lack of visual clarity). A virtual body rendered directly on the screens (bypassing the resolution limits of the VR cameras) reaches a 1600x1440 per-eye resolution.

Therefore, the score is not printed but is visualized inside the AR environment as a virtual body (virtual score). More precisely, each page of the score is used as the texture of a virtual rectangle (plane). Every time a page is turned, the next page's picture is loaded as the new texture. Pages can be turned by using a dedicated interface near to the score.

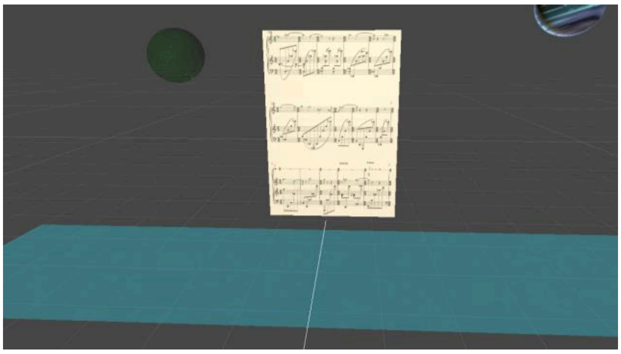

Figure 8. The score in the Unity Editor

4. DISCUSSION

The presented interfaces for piano augmentation have been developed as a compositional resource for this specific project, not as a tool for being used by other composers in a more general capacity. From the author's point of view, this is not a limit, but rather a particularity of AR applications for musical performance: designing an AR interface is relatively time-inexpensive compared to the construction of a physical one, where, beyond the software development, the assembly of hardware components comes into play. AR technology provides a fast and intuitive way to design interfaces, making it easier to conceive them as a component of the compositional process (i.e., a set of interfaces is created for only one composition and designed depending on the specific musical context).

The clear visual presence in space makes any form of augmentation more intuitive to approach for a performer, who does not necessarily have to memorize positions in space or specific movements. Additionally, if provided with haptic wearable devices, virtual bodies can be associated with a tactile feedback, thus enhancing immersivity.

A major advantage of a virtual AR interface over a real one consists in its possibility to change over time. Changes might include functions, shape, position in space, appearing and disappearing. Especially for composite interfaces, the arrangement of triggers and sliders can be changed over time, according to the most comfortable position for playing both the piano keys and the virtual objects. Again, the visual presence of such objects makes it easier to provide indications for the interpreter and allows more articulated designs strategies without increasing the steepness of the learning curve for the performer. In fact, while a similar set of possibilities could be available by just using sensors and software implementations (with no AR components), the player would need to memorize the different functions and actions/positions of interaction, with no obvious indication in time and space.

With the use of interface-notation hybrids, the gestural components of the composition can be designed and choregraphed and the gesture itself finds a new possibility of expression.

All the presented interfaces are relatively simple to use and do not include advanced functionalities in terms of gesture recognition and analysis of the physical movement (e.g., linking specific shapes of the gesture to some specific dynamic/spectral outcome). In this sense, the kind of augmentation presented here is not as developed as a fully-fledged Digital Instrument or Hyperinstrument could be. However, articulated practices can be recovered through a high accuracy in spatial interaction: thanks to the visual feedback, the composer can require to the performer to narrowly differentiate interaction between close points in space.

The mirroring to a projector of the point of view of the player poses interesting questions about the interface design from the audience perspective. In fact, the visual dimension itself becomes an element to develop in time and to compose, along with the interaction and its graphic feedback. Interfaces for instrumental augmentation are not only important for their functional capacity but can also be considered and designed as a visual component of a multimedia composition.

In some future work, an evaluation of the composition with several performers could be valuable for developing more effective design strategies.

5. CONCLUSIONS

This article has presented some possibilities of design for interfaces in AR, meant as a form of instrumental augmentation.

The mentioned interfaces can be of four kinds: object-triggers, multi-dimensional sliders, composite interfaces, interface-notation hybrids.

The visual presence in space of such augmenting virtual devices allows a better-than-before space resolution for mapping position-dependent functions without requiring any memorization process. The precision of the space sampling allows more virtuosic ways of interaction and the indication of gestures to perform in 3 dimensions. The resulting piano augmentation makes the input points look and behave like a physical extension of the piano.

The flexibility of design in AR is one of the most promising characteristics: interfaces can be modified over time (in position, appearance and function), and such changes can be made immediately intuitive.

The need to also consider the visual quality of AR in music performance from the audience's point of view potentially has a deep impact on the work of a composer making use of such a technology. In fact, interactions, visual feedbacks, interfaces behavior and gestural quality of virtual bodies also have the potential to carry an expressive visual component which can be developed and composed in time as much as the sonic dimension.

6. REFERENCES

[1] B. Frédéric, N. Rasamimanana, E. Fléty, S. Lemouton, and F. Baschet, "The Augmented Violin Project: reseach, composition and performance report," *Proc. 2006 Int. Conf. New Interfaces Music. Expr. (NIME), Paris, Fr.*, 2006.

[2] T. Machover, "Hyperinstruments - a progress report 1987-1991," 1992.

[3] J. Malloch, D. Birnbaum, E. Sinyor, and M. M. Wanderley, "Towards a new conceptual framework for digital musical instruments," in *Proceedings of the 9th International Conference on Digital Audio Effects*, 2006.

[4] S. Serafin, C. Erkut, J. Kojs, N. C. Nilsson, and R. Nordahl, "Virtual Reality Musical Instruments: State of the Art, Design Principles, and Future Directions," *Comput. Music J.*, 2016.

[5] J. Fillwalk, "ChromaChord: A virtual musical instrument," in *2015 IEEE Symposium on 3D User Interfaces, 3DUI 2015 - Proceedings*, 2015.

[6] G. Santini, "Synesthesizer: Physical Modelling and Machine Learning for a Color-Based Synthesizer in Virtual Reality," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019.

[7] I. Hwang, H. Son, and J. R. Kim, "AirPiano: Enhancing music playing experience in virtual reality with mid-air haptic feedback," in *2017 IEEE World Haptics Conference, WHC 2017*, 2017.

[8] D. Johnson and G. Tzanetakis, "VRMin: Using Mixed Reality to Augment the Theremin for Musical Tutoring," *NIME 2017 Proc. Int. Conf. New Interfaces Music. Expr.*, 2017.

[9] M. Weing *et al.*, "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.

[10] X. Xiao, L. H. Hantrakul, and H. Ishii, "MirrorFugue for the Composer, Performer and Improviser," 2016.

[11] F. Trujano, M. Khan, and P. Maes, "Arpiano efficient music learning using augmented reality," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.

[12] A. Birhanu and S. Rank, "KeynVision: Exploring piano pedagogy in mixed reality," in *CHI PLAY 2017 Extended Abstracts - Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, 2017.

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

[14] G. Santini, "Composing space in the space: an Augmented and Virtual Reality sound spatialization system," in *Sound and Music Computing 2019*, 2019, pp. 229–233.

[15] 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.

[16] ARShow, "ARShow," 2020. [Online]. Available: http://www.arshowpro.com/#services. [Accessed: 18-Jan-2020].

[17] S. Kelly *et al.*, "AR-ia: Volumetric opera for mobile augmented reality," in *SIGGRAPH Asia 2019 XR, SA 2019*, 2019.

7. APPENDIX

Visualize supporting material at https://www.giovannisantini.com/studisullarealtanuova