Abstract Reviewed Paper at ICSA 2019

Presented * by VDT.

Mixed Realities: a live collaborative musical performance

A. Genovese, M. Gospodarek, A. Roginska
New York University, USA, Email: genovese@nyu.edu

Abstract

In the presented work, a live-rendered percussionist is transformed into a virtual game character and performs a piece along digital avatars created from recordings (audio and motion-capture) of other members of an ensemble, while audience members can observe the collaborative performance through VR headsets. To create a cohesive and compelling result, the auditory expectations of the listeners need to be considered in terms of acoustic integrity between real and virtual sources and spatial impression of each avatar performer.

This paper presents an overview of the workflow and motivations behind this pilot experiment of a novel musical experience, laying down the foundations for future subjective studies into collaborative music performances using virtual and augmented reality headsets. Particular focus is given to the technical challenges concerning the audio material, the perspectives of each participant role, and the qualitative impressions of musicians and audience.

1. Introduction

The question of digitally augmented music collaboration is ever-relevant in the field of immersive audio and musical performance. Currently, the adoption of commercial immersive headset devices for virtual and mixed reality can enable today's musicians to experience enhanced forms of virtual presence when connected to their peers. For instance, motion-capture data of a human being can be streamed and live-rendered using complex camera-sensor systems. Game avatars animated by real human beings are thus brought into a shared virtual space where observers and participants are part of a common world. A performer's captured audio stream can be paired with associated metadata to create a three-dimensional trajectory of an audio source, or to create an association to a specific digital element, or avatar, visualizable in the shared virtual scene via a head-mounted display device (HMD, or headset). Usually, the main goal of this application is to create an illusion of realism (or hyper-realism) for the participants involved while maintaining an effective musical output quality. In other words, to achieve a sensation where the collaborating musicians are perceived, by themselves or by an audience, as if playing together in the same room.
Such prototype systems have been recently explored for immersive theatre applications where live or recorded actors are experienced virtually by a co-located audience [1, 2]. Musical environments have also been explored by testing designs of musical interfaces [3, 4] or assessing the factors that can improve the perceived experience plausibility [5] and co-presence [6, 7]. However, music-making and concert experiences in the traditional sense have so far been more complex challenges, not extensively covered in literature.

As these technologies become progressively more available, it is worth to explore the implications for new kinds of enhanced music performances. Using HMDs, musicians and audience members alike can be placed in a shared virtual world while being, for instance, in physical remote locations. Another avenue is given by Mixed Reality (MR) scenarios, where digital elements are rendered with a degree of connection to the physical world of a user [8]. MR is interesting as it raises a question of considering the individual perspective of each participant in terms of the features that define their local reality both visually and acoustically. In MR applications,

* Please note that the papers at ICSA can be published by VDT, in print, online and as PDF download.

the real and the virtual elements are treated to blend together into a cohesive scene creating an illusion of an augmented world where virtual objects assume location-specific properties, thus seeming more ”realistic”. In terms of audio, this involves acoustic character of the space, (early reflections, late reverberation, etc.), characteristics of the audio source (e.g. radiation patterns) and spatialized, dynamic, 3D localization of individual sources.
The number of variables and possible combinations involved into organizing immersive mixed-reality performances is only bounded by the amount of technology at the disposal of the artists and engineers. Since it is a complex, if not impossible, task to generalize a qualitative study for all kind of virtual music network topologies, it is usually necessary to limit the investigation by having a target user and scenario in mind.

We hereby present a pilot implementation of a novel mixed-reality musical experience using motion-capture and VR headsets, meant to serve as a first test and canvas upon which to build a series of formal studies on the subject. Although only non-formal assessment was conducted for this work, the establishment of a workflow pipeline helped to investigate an exploratory scenario and provide insights on future evaluation methods. This paper discusses the design, implementation and technical challenges of this work, with particular focus on the audio elements.
In the presented work, a dancer and three percussionists of an African music ensemble have been individually recorded with microphones and an OptiTrack motion capture system. Their audiovisual data was then converted into digital avatars able to be reproduced as virtual performer animations through VR headsets. During an exhibition demo, an additional live percussionist, wearing a motion-capture suit, was brought into the virtual scene as a real-time character. A single audience member, located in the same space as the live-percussionist, was able to observe the joint collaborative performance of all musician avatars, live and pre-recorded, by means of a VR headset and headphones.

2. Experience Design

For this particular piece, it was decided to explore a mixed-reality setting where a combination of live and pre-recorded musician game avatars (created with audio and motion capture data) would play together and be observable within a VR headset by an audience member. Within this framework, a number of possibilities were open for exploration in terms of organization and disposition of the three types of participants involved (audience, live musician and pre-recorded musicians, from here referred to as the ”virtual ensemble”). What was of interest was the question of how to design a convincing sense of ”co-presence” and how to create a sense of shared reality from the perspective of the target user, in this case, the audience. To achieve this, it was decided to tailor a mixed-reality (MR) experience grounded in the local auditory reality of the audience. In other words, fit the virtual material to adapt to a predefined ”concert” space shared by the both the live performer and

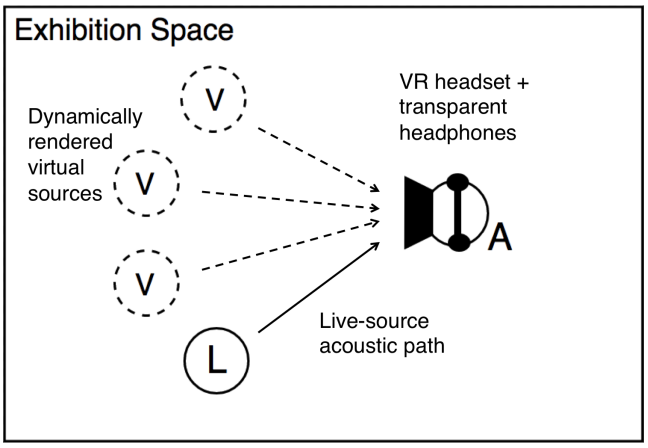

Fig. 1: Design for the spatial disposition of participants during the exhibition phase. The live musician (L) shares the room with the audience (A). The virtual avatars (V) are spatially located around the listener to form a virtual ensemble with the live performer. The audience is provided with transparent headphones in order to allow the local acoustic path to be heard with little obstruction while dynamically rendering in binaural format the sound of the virtual members (acoustically treated).

audience participants. Within this shared space, it was desired to make the experience ”feel like a concert” where all the musicians could be perceived as if ”being together in the same space” [9]. Using this *audience-centric* approach, all the elements had to be put into service for the enhancement of the audience perspective and quality of experience, but also without compromising the performance of the musician.

Mixed-reality experiences have a different set of challenges in comparison to pure virtual reality experiences. Having a sonic reference from the real world allows the auditory system to compare the real and virtual cues and base the judgment of ”scene realism”, or plausibility, upon the correlation between the two. In this scenario, as the live musician shares the concert space with the listener, the sound emitted by the ”local” instrument and its natural acoustic room reflections, is heard by the audience. This forms the perceived ground reality to which the ”remote” virtual sound needs to adapt towards. For the virtual content to mix and adapt cohesively to the concert space and the local acoustic character, artificial reverberation using a room-impulse-response (RIR) measured *in-loco* can be overlayed through convolution to each object source, provided the source-material is recorded in dry condition. This step ensures a shared acoustic character to all elements, helping to build the auditory integration of the scene [10], while small mismatches of that character could instead potentially decrease that sense of integration.

From the perspective of the audience, the live sound needs to be heard naturally as it fits with the visual display, whether real or virtual, thus unobstructed by closed surfaces between the ears and the direct path of the source. In regards to the virtual sources, a dynamic binaural rendering process can create externalized, localizable, virtual sound emitters, which help the listener to build a cognitive organization of the stage display. These considerations can be addressed by the use of

	VE	LM	Aud.
Auditory Env.	Virt.	Real	Real & Virt.
Visual Env.	Virt.	Real	Virt.

Tab. 1: Table of shared auditory and visual environments between the three participants: virtual ensemble (VE), live musician (LM), and audience (Aud.). The labels indicate "Real" environment (Re.) or "Virtual" environment (Virt.).

open-back headphones connected to a binaural engine and a head-tracking device. Open-back headphones allows for local real sound sources to be perceived reasonably uncolored, while the integrated IMU units in modern HMDs can provide the positional data for six-degrees-of-freedom movements of a user. In fact, when the 3D binaural rendering is made responsive to rotations and position shifts it dramatically enhances the spatial impression, presence, and perceived directional accuracy [11]. Ideally, a transparent, or hear-through headphone device would be better than the open-back kind as it provides the least amount of coloration of free-field sources [12]. The use of a loudspeaker system is theoretically also possible but it makes it harder to accurately reproduce the directivity patterns of the instruments.

The exhibition of the performance was thus planned as portrayed in Fig. 1. The local presence of a real sound source in proximity to the listener, in conjunction to the delivery of spatialized object-based audio via headphones, effectively creates an auditory mixed-reality scenario on top of which, a virtual visual environment has to be projected on a VR HMD. The resulting dynamic is that of a combination of shared environments between the three types of participants (Tab. 1). Most interestingly, the audience shares the auditory space with the live musician, but the visual space of the virtual ensemble (including a live-rendition of the live musician). To achieve congruence in any realistic display, the bond between the visual and auditory senses needs to be considered; a listener's expectation of the acoustic character of a space is in fact influenced by its visual impact [13]. Cohesiveness between the two sensorial realms is key for achieving a convincing mixed-reality base of display and avoid poor engagement with the intended application. In practice, as a VR headset is here needed to display the avatars and provide the head-tracking data, the virtual visual environment projected to a viewer needs to be congruent with their experience of the local auditory reality if a compelling blend is desired between the live drummer and the virtual ensemble. As we grounded this experience based on the local sound, the visuals need to adapt by displaying an environmental location which would plausibly relate to the local acoustics.

3. Implementation

The implementation of this experience can be divided into three main phases, data capturing, scene design, and exhibition. While the capturing phase was conducted to collect material also usable for possible independent projects, the material was selected to be specifically appropriate for the exhibition described in this paper.

The selection of musical content for the performance had to take into consideration the use of motion capture sensors suits which may interfere in the regular use of an instrument. The selection of an African percussion quartet (Djembe), plus a dancer, was deemed appropriate as the suit proved to be non-invasive to the musicians and also helped to later simplify the graphical rendition of the instrument in the virtual scene. Percussive, highly-transient, content is also more easily localizable in binaural displays as its wide frequency bandwidth covers the range of both ITD and IID cues in generic HRTF sets [14]. The presence of the dancer was effective in raising the aesthetic value of the final piece but was not a determinant element of the exhibition presented.

3.1. Motion and audio capture

The capturing session was performed in a medium size room equipped with optical cameras for motion tracking. Each of the musicians was recorded separately to ensure full control over each take. The goal was to obtain individual, dry recordings of each member of the ensemble so that each one could be treated as a separate audiovisual object.

The motion tracking was performed using an OptiTrack system with 15 cameras and the Motive capture software [15]. Each percussionist, and the dancer, were fitted with a 32-sensors suit, including gloves. As the suit did not seem to impair the musical performance ability, the audio recording could happened simultaneously to the motion-tracking.
For recording the sound of the Djembe drum, four condenser microphones with cardioid pattern were used. This directivity pattern was chosen in order to achieve a good separation of the direct sound to the recording room reflections. The top two microphones were oriented towards the drum membrane, while the bottom two were placed close to the resonance chamber. The ambience sound was also captured using the first order Ambisonics microphone Sennheiser Ambeo, positioned in front of the drum. The ambience sound was only collected for reference and future use, and not included in later parts of this production.

The main drummer was recorded first using a metronome click. The best take was chosen and the track was used as a reference for the following drummers who were recording the other voices of the percussion piece. In total, three drumming parts and a dancer part were recorded, leaving the last drummer part for the live exhibition.
Since it is common, when capturing full body motions, that sporadic glitches may occur for some of the sensors, the motion-tracking data had to be passed through a "cleaning" procedure before further editing. This consisted in correcting the position of each individual sensors at the frames where those glitches resulted in an unnatural skeleton rendering.

3.2. VR Scene Design

The NYU Future Reality Lab provided the necessary facilities that the project required, a large enough room for the performance needs, and a motion capture system for the live-

musician. All the steps described in this section involved a tailored approach to this particular space.

3.2.1. Visual Design

An additional important reason for the selection of the room, was the availability of a 3D model rendition of the actual performance space as a digital asset. This made it possible to provide to the audience a semi-identical virtual environment to the actual physical one, thus optimizing the coherence of the local sound with the visual stimulus.
The VR scene was built using the Unity game engine [16] using the aforementioned laboratory asset and digital models of African percussion instruments. The cleaned and synced motion-capture data was used to create skeleton rigs in the AutoDesk Maya software [17] which were used to animate digital characters back in the Unity scene. The three virtual percussionists were disposed as shown in Fig. 1 while the dancer figure was placed in the background behind the digital ensemble.
An Optitrack tracking system was mounted in order to capture the live musician during the exhibition and transform the motion into an additional character in Unity (live-streaming was implemented through the Motive asset package). A calibration step was performed in order to achieve a one-to-one spatial match between the digital rendering of the live musician (plus a digital drum model) and its actual physical position in the room. This was important in order to ensure the perfect match between localization of visual avatar and the live free-field sound of the performer.

3.2.2. Audio rendering

To obtain an acoustic characterization of the space, measurements of the room acoustic impulse response were retrieved from an earlier project conducted in the same location. Two omnidirectional microphones DPA 4006 were mounted in the center of the performance space 17 cm apart and at the height of 1.8 m. The measurement signal was reproduced by one speaker located 3 m from the microphones. The ScanIR (v2) MATLAB toolbox [18] was utilized to reproduce the measurement signal and capture the impulse responses. A 2-seconds sinesweep was thus recorded in stereo at 96 kHz, ranging from 20 Hz to 20 kHz [19]. The impulse responses were later used to process the signal and superimpose the acoustic characteristics of the exhibition space onto the rendered audio tracks, increasing the timbral blend between performers.

For each of the musicians, two microphone positions, one from the bottom and one from the top of Djembe drum were selected out of the four available, and rendered as separate audio tracks. A gentle compression was applied to achieve a more satisfactory timbre of the instrument. The tracks were used to create virtual audio objects implemented in Unity within the prepared visual scene. The Steam-Audio plugin [20] was employed as the audio rendering engine. Each virtual Djembe model was assigned two object sound sources, one for each of the two top-bottom microphone tracks. The transient slap sound from the hand hitting the membrane was thus placed at the top of each digital instrument while the low frequency resonance was positioned at the bottom. This double-emitter strategy ensured that the size of the instruments and their radiation characteristics were preserved.

The scene spatialization was handled by the binaural rendering engine of Steam Audio, using generic HRTFs and the rotation/position data from the VR headset IMU unit and its tracking sensors, which allow for dynamic perspective. The location of each of the object sound sources was rendered according to its avatar position in relation to the viewer, while distance attenuation was simulated by use of the square-law [21].
In order to apply the diffused reverberation of the room, all the microphone tracks were summed together and the resulted signal was convolved with the late diffused part of the stereo impulse response earlier collected. Finally, the resulting reverberant stereo file was mixed with the dry binaural mix of the avatars (where a slight EQ was applied), in order to achieve a calibrated balance of direct vs reverberant sound deemed aesthetically satisfactory for a compelling experience.

3.3. Exhibition

The experience demo was exhibited to a crowd of academics during an internal event. Rehearsals with the live musician were first conducted in order to test the system and allow sufficient comfort in performing while wearing a tracking suit (Fig. 2). It is worth to note that the musician was also part of the original motion-captured ensemble, meaning that there was familiarity with all the parts in the piece and with performing with the suit. The live performer was not provided with a separate VR headset, but with the binaural audio stream deriving from the movements of the audience device. This was not ideal in terms of providing the musician with an optimal auditory perspective but ensured that a perfect synchronization with the recorded material could be maintained. Because a single room impulse response was available, the audio levels had to be mixed for a single seating location of the audience in the space. The audience seat was placed in front of the virtual ensemble and the balance between direct

Fig. 2: Video still of the exhibition rehearsal. The VR audience POV of the audience is shown in the background picture, while the overlayed smaller picture illustrates the external view of the live musician and the audience, seen from the experimenter. A video of the exhibition rehearsal is available at https://www.youtube.com/watch?v=-0VqIn1pTA0.

and reverberation sound was adjusted for that location. The audience (one person at a time) was fitted with a tethered VR headset and open-back headphones, and encouraged to look around and shift their head position within the area of their seat, but to not walk around the performance space. No formal questionnaire data was collected from the audience, but the musician participants had a chance to share their impressions with the authors.

4. Discussion and future work

The work described in this paper was exhibited in an informal setting closer to an art show rather than a formalized experiment. However, a lot of insights were gathered as well as the consolidation of the design criteria that lead to this kind of implementations.

Although the design was audience-centric, no particular audience feedback was received other than general comments about the visual quality of the meshes. However, the live performer was able to respond to a short post-performance questionnaire made of three scale ratings and an open-ended feedback form. The musician rated the comfort of performing while wearing the suit as 3 out of 7 (1 being "Very uncomfortable" and 7 being "Very comfortable"). The sense of acoustic cohesiveness between real and virtual sound was rated 5 out 7 (1 being "Not cohesive at all" and 7 as "Very cohesive"). Finally, the difficulty of playing with avatars rather than real life musician was rated to be 3 on a seven-point likert scale (1 being "Easier", 7 being "Harder", with 4 being "No Difference"). Other general impressions included the fact that the performance felt like a "one-way avenue of communication, where my job was to fit myself into this world that was created for the experience" and it "did not feel as organic as performing with other people in real time, perhaps because of some visual aspects". Naturally, this is a single data point which needs to be further explored before generalizing to wider settings.

These answers will be taken into consideration for future work, and compared to possible situations where instead all the participants will be live. Some improvements might be achieved in future by treating the live musician point-of-view with the same approach as the audience perspective (acoustic adaptation and binaural dynamic response from their own perspective). Our project also assumed that the audience members were not changing their position drastically since the implemented impulse response was captured at a single point in space with an omnidirectional mic. In future implementations, adding more measurements and dynamic controls for the direct sound to reverberation ratio could add a further degree of realism to the scene and give more flexibility of movement for the audience. More accurate room impulse responses would be obtained by adjusting the source-emitter position at the intended locations of each participant, adding a more accurate early-reflection pattern rather than just the diffused part.

Having built this pilot framework, future studies on mixed-reality and music performance will be conducted in order to investigate the technical and cognitive aspects which regulate the subjective quality of experience. When talking about evaluating and measuring such quality of experience, it is important to differentiate the metric according to the role of a participant. A setting can be indeed defined in terms of the *target-user*, indicating if the outcome needs to be compelling to the musician for a music-making experience, or to an audience for a concert. The perspectives of the audience and the live-musician could be both taken into account into deciphering the success of musical virtual environments. While they both might tie quality to their sense of presence into the scene, the musician might seek something more keen to an intersection of "co-presence" and "naturalness", as in the sense of "being performing together" to the fellow performer, in a setting comparable to real life. The evaluations from the two perspectives might or might not correlate. However, some kind of different design choices oriented towards *hyper-realism* or *unrealism* might have to be evaluated not in terms of naturalness (or realism), but in terms of telepresence and plausibility, which connect respectively more with a general sense of engagement and coherence between not-necessarily-real environments.

Qualitative observations through questionnaires might reveal how some of these subjective attributes vary according to the nature of the content and the network topology created (e.g. distribution of musicians between local and "remote", or live and virtual). Furthermore, some alternative strategies such as task-success metrics (e.g. correctness of musical output, musical synchronization ability, etc.) or parametric control by the audience [5], could provide a more objective measure and reveal the relationship between technical aspects, perceived quality and effective outcome.

The outlook of future applications related to this work its oriented towards studying virtual and mixed reality for distributed music networks. Music collaboration over the internet has been explored for years [22] but only now its possible to explore virtual performance spaces and connect these endeavours with a three-dimensional virtual presence of the musical performers. In this field, latency is still the primary concern, and streaming motion-capture data efficiently over the internet is a challenge yet to be overcome. However, putting this issue aside, it is still worth to study all the other aspects which relate to these systems.

In our case, the use of pre-recorded material for the guest musician to perform a part, can allow a test system to bypass the issues of signal latency that are inherent in music network collaborations. This also helped to focus the effort on the experience design and reduce rehearsal times for the musicians involved. In future iterations of this experience, it is intended to simulate distributed settings where all musicians are live in order to fully study a real-time collaborative mixed-reality music environment in either musician-centric or audience-centric evaluation schemes.

5. Acknowledgments

The authors would like to thank all people who helped in the building of this demo. Christopher Allen O'Leary (drummer and live performer), Max Meyer (drummer and dancer), Jared

Shaw(drummer), Sripathi Sidhar, Christy Welch, Scott Murakami (audio engineers), Robert Pahle (IT support), Pasan Dharmasena (rendering of avatars). We would like to thank also the Future Reality Lab at NYU for lending the space and technology.

6. References

[1] Kris Layng, Ken Perlin, Sebastian Herscher, Corinne Brenner, and Thomas Meduri, "Cave: Making collective virtual narrative: Best paper award," *Leonardo*, vol. 52, no. 4, pp. 349–356, 2019.

[2] David Gochfeld, Corinne Brenner, Kris Layng, Sebastian Herscher, Connor DeFanti, Marta Olko, David Shinn, Stephanie Riggs, Clara Fernandez-Vara, and Ken Perlin, "Holojam in wonderland: Immersive mixed reality theater," *Leonardo*, vol. 51, no. 4, pp. 362–367, 2018.

[3] Thomas Deacon, Tony Stockman, and Mathieu Barthet, "User experience in an interactive music virtual reality system: An exploratory study," in *Bridging People and Sound*, Mitsuko Aramaki, Richard Kronland-Martinet, and Sølvi Ystad, Eds., Cham, 2017, pp. 192–216, Springer International Publishing.

[4] Stefania Serafin, Cumhur Erkut, Juraj Kojs, Niels C. Nilsson, and Rolf Nordahl, "Virtual reality musical instruments: State of the art, design principles, and future directions," *Computer Music Journal*, vol. 40, no. 3, pp. 22–40, 2016.

[5] I. Bergström, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater, "The plausibility of a string quartet performance in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1352–1359, April 2017.

[6] Byungdae Jung, Jaein Hwang, Sangyoon Lee, Gerard Jounghyun Kim, and Hyunbin Kim, "Incorporating co-presence in distributed virtual music environment," in *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM, 2000, pp. 206–211.

[7] Pontus Larsson, Aleksander Väljamäe, Daniel Västfjäll, Ana Tajadura-jiménez, and Mendel Kleiner, "Auditory-Induced Presence in Mixed Reality Environments and Related Technology," in *The Engineering of Mixed Reality Systems*, chapter 8, pp. 143–163. Springer-Verlag, 2009.

[8] Ina Wagner, Rod Mccall, Ann Morrison, and Marne Valle, "On the Role of Presence in Mixed Reality," *Presence*, vol. 18, no. 4, pp. 249–276, 2009.

[9] Saniye Tugba Bulu, "Place presence, social presence, co-presence, and satisfaction in virtual worlds," *Computers & Education*, vol. 58, no. 1, pp. 154–161, 2012.

[10] Will Bailey and Bruno M. Fazenda, "The effect of visual cues and binaural rendering method on plausibility in virtual environments," in *Proceedings of the 144th AES Convention*, Milan, Italy, 2018.

[11] Durand R. Begault, Elizabeth M. Wenzel, and Mark R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source.," *Journal of the Audio Engineering Society. Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[12] Stefan Liebich, Raphael Brandis, Johannes Fabry, Peter Jax, and Peter Vary, "Active occlusion cancellation with hear-through equalization for headphones," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 241–245.

[13] Daniel L Valente and Jonas Braasch, "Subjective scaling of spatial room acoustic parameters influenced by visual environmental cues," *The Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 1952–1964, 2010.

[14] Jens Blauert, *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

[15] "Motive - optical motion capture software.," https://optitrack.com/products/motive/, (Accessed on 08/27/2019).

[16] "Unity real-time development platform — 3d, 2d vr & ar visualizations," https://unity.com/, (Accessed on 08/01/2019).

[17] "Autodesk maya — 3d computer animation, modeling, simulation, and rendering software," https://www.autodesk.com/products/maya/overview, (Accessed on 08/27/2019).

[18] Julian Vanasse, Andrea Genovese, and Agnieszka Roginska, "Multichannel impulse response measurements in matlab: An update on scanir," in *Proceedings of the AES International Conference on Immersive and Interactive Audio*, York, UK, 2019.

[19] Angelo Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[20] "Steam audio," https://valvesoftware.github.io/steam-audio/, (Accessed on 08/01/2019).

[21] BG Shinn-Cunningham, "Distance cues for virtual auditory space," *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia*, pp. 227–230, 2000.

[22] Cristina Rottondi, Chris Chafe, Claudio Allocchio, and Augusto Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.