Immersive Orchestras: audio processing for orchestral music VR content

Jordi Janer, Emilia Gómez, Agustín Martorell and Marius Miron
Music Technology Group
Universitat Pompeu Fabra
Barcelona
Email: first.last@upf.edu

Benjamin de Wit
DutchDays VR
Amsterdam
Email: info@dutchvrdays.nl

Abstract—This paper combines Audio Signal Processing and Virtual Reality (VR) content to create novel immersive experiences for orchestral music audiences. In VR, the auralization of sound sources of recorded live content remains still a rather unexplored topic. We aim to build a multimodal experience, where visual and audio cues bring a sonic augmentation of the real scene. In the particular scenario of orchestral music content, our goal is to acoustically zoom on a particular instrument when the VR user stares at it. This work aims to improve the learning aspects of music listening, either for education or for personal enrichment. We use audio signal processing to separate different sound sources (instruments) in a acoustic scene (orchestral music recording). Given the signals captured by multiple microphones and the musical score of the piece, our system is able to isolate the different instruments. From the processed separated tracks, we use a binaural rendering technique to emphasize a give instrument. For these experiments we used original content from top European orchestras.

I. INTRODUCTION

This work puts together the outcomes of two existing projects around classical music and new technologies. On the one hand, the Music Technology Group (Universitat Pompeu Fabra, Barcelona) participated in the EU H2020 Phenicx project (2013-2016) , which aimed at enhancing the experience in orchestral music concerts though music analysis and audio processing technologies. On the other hand, we count on the collaboration of Benjamin de Wit who worked as a creative VR producer with WeMakeVR (Amsterdam), a VR technology Company that records, produces and distributes VR content of Orchestral music performances. WeMakeVR has recorded prestigious Orchestras including the Berliner Philarmoniker and the London Symphony Orchestra.

The introduction of new media technologies in the world of Classical Music is receiving more and more attention. Orchestras around the globe start publishing their program notes as iPad apps, and offer digital streaming of their performances in real-time. How can we design new interactive applications that use Virtual Reality (VR) in the context Classical Music?

In this paper we present a proof-of-concept that uses VR content and audio processing research to create acoustic experiences that go beyond the visualization of 360 video recording. The user has an immersive multimodal experience where audio and visual cues are synchronized to generate an *acoustic zoom* effect. We employ Audio Source Separation to individually isolate the orchestra instruments. In the VR rendering, the users front view controls the presence of the instrumental source to create this effect of Acoustic Zoom. We prepared examples of VR recordings of Berliner Philarmoniker and the London Symphony Orchestra performances, and the audio source separation results obtained from the Phenicx EU research project.

Our goal is to blend in this presentation art (music), technology (VR), past (classical repertoire) and future (new user applications). We believe music learning both for education or personal enrichment can benefit from advances in the area of Virtual and Serious Games.

II. AUDIO SOURCE SEPARATION AND INSTRUMENT EMPHASIS

Our Audio Source Separation method [1] allows to process multi-track orchestral recordings to obtain instrument separated tracks. Once we have the separated tracks, we can emphasize a particular instrument over the full orchestra downmix recording. We call this process *Instrument Emphasis*. Our workflow consists in processing multi-channel orchestral recording from which the musical score is available. The multi-track recordings are obtained from a typical on-stage setup in a concert hall, where multiple microphones are placed on stage at certain distance of the sources. The goal is to reduce leakage of other sections, obtaining a *clean* signal for the selected instrument.

In terms of system integration this application has two parts. The Front-end, as shown in Figure 1 is responsible to interact with the content producer in the uploading of media content and present the results to the user. The back-end is responsible to manage the audio data workflow between the different signal processing components. We process the audio files in batch estimating the signal decomposition for the full-length. For long audio files, as in the case of symphonic recordings, the memory requirements can be too demanding even for a server infrastructure. Therefore, to overcome this limitation, audio files are split into blocks. After the separation has been performed, the blocks associated to each instrument are concatenated, resulting in the separated tracks. The separation quality is not degraded if the blocks have sufficient duration. In our case we set the block duration to 1 second.
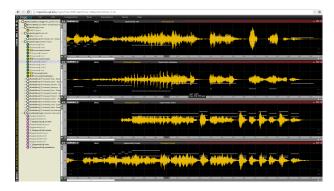
Fig. 1. Screen shot of the online production interface to generate the Instrument Emphasis tracks.

Fig. 2. Example of VR content produced by Los Angeles Philarmonic publicly available on Youtube[4].

It ensures processing an audio file long enough to disregard these limitations. Example of this application can be found online as a web demo, and in this video.

III. AUGMENTING ORCHESTRAL CONTENT IN VR

Orchestras around the world are already producing experimental VR productions[1].The audio possibilities offered to the user in these applications are still limited. Our approach uses the processed audio material to augment the acoustic scene, controlling dynamically the spatialization of the musical sources (instruments). In this case the auralization aims at recreating acoustically the recorded performance from a specific listening location and orientation, with a controllable disposition of instruments on the stage and the listener.

We have considered binaural synthesis as the most suitable spatial audio technique for this application. Humans locate the direction of incoming sound based on a number of cues: depending on the angle and distance between listener and source, the sound will arrive with a different intensity and at different time instances at both ears. The idea behind binaural synthesis is to artificially generate these cues to be able to create an illusion of directivity of a sound source when reproduced over headphones [2], [3]. For a rapid integration into the virtual reality prototype we have used the non-commercial plugin for Unity3D of binaural synthesis provided by 3DCeption [2].

Specifically, these type of applications open news possibilities in the area of Virtual Reality (VR), where video companies are already producing orchestral performances specifically recorded for VR experiences as showed in Figure 2.

Using a VR headset with headphones, the Instrument Emphasis application is able to perform and acoustic zoom effect when pointing at a given instrument or section, knowing the orchestra layout and the camera position.

IV. USE CASES

Principally the motivation behind this work is to provide to the classical music students, professionals and aficionados new ways of experiencing the music. Compared to other musical genres, classical music audiences are less heterogeneous, both in age and sociocultural position. This audience is often coined as conservative, where innovative applications such as VR will hardly get adopted. With this work we strive to demonstrate its interest in at least two particular use-cases:

- Learning: instrument performance or musicology students can benefit from the possibility to listen to specific instruments, analyzing performance aspects that are often masked by the rest of the orchestra.
- Personal Enrichment: classical music aficionados are typically eager to discover all details of their preferred repertoire. The possibility of focusing the aural attention in specific instruments during a performance is something not possible in a concert hall that VR can provide.

V. CONCLUSION

This work introduces a collaboration between one academic and one industrial partners in the areas of audio signal processing and VR technologies. Using state of the art audio processing methods, we show how to augment the auralization of orchestral music in VR recordings. Our approach, called Instrument Emphasis, allows to acoustically zooming in the instrument the user is facing. Departing from the current proof-of-concept, further research work will involve formal user evaluations to determine the auralization preferences from a user standpoint.

ACKNOWLEDGMENT

REFERENCES

[1] M. Miron, J. J. Carabias, and J. Janer, "Improving score-informed source separation for classical music through note refinement," in *16th International Society for Music Information Retrieval (ISMIR) Conference.*, Malaga, 2015.

[2] D. R. Begault and E. M. Wenzel, "Headphone localization of speech," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 35, no. 2, pp. 361–376, 1993.

[3] G. S. Kendall, "A 3-d sound primer: directional hearing and stereo reproduction," *Computer music journal*, vol. 19, no. 4, pp. 23–46, 1995.

[4] "Orchetra vr, los angeles philarmonic," https://www.youtube.com/watch?v=j-vtJYmia_o.

[1]Orchetra VR, Los Angeles Philarmonic, http://www.laphil.com

[2]https://twobigears.com/index.php