

# 3D AUDIO AR/VR CAPTURE AND REPRODUCTION SETUP FOR AURALIZATION OF SOUNDSCAPES

Rishabh Gupta, Bhan Lam, Joo-Young Hong, Zhen-Ting Ong, Woon-Seng Gan, Shyh Hao Chong, Jing Feng

*Nanyang Technological University, School of Electrical and Electronic Engineering, Digital Signal Processing Lab, Singapore*

*email: [grishabh@ntu.edu.sg](mailto:grishabh@ntu.edu.sg)*

Auralization technique is often used to reproduce sound signals and model acoustic transmission effects in a real environment. This paper describes the construction of an omnidirectional video and ambisonic audio capturing device for Virtual Reality (VR) reproduction, which is intended to be used for auralizing soundscapes. The audio-visual stimulus projected in the VR headset depicts a “360°” scene of a noisy park. The rendering tools and methodology has been described. Different positive soundtracks were analysed and rendered at different spatial locations. An Augmented Reality device (AR) coupled with open back headphones was designed to simulate the masking effects. The aim was to study the effects of varied spatial position of the masker on the static noise source. The devices and tools described can be used in various applications like studying level difference for different types of masking, conducting different types of audio only or audio-visual experiments and for evaluating the effectiveness of placing a virtual sound object in different spatial orientations with respect to the noise source.

Keywords: Virtual Reality, Auralization, Ambisonics, 3D audio, Augmented Reality

---

## 1. Introduction

Auralization is defined by Vorländer [1] as “the technique of creating audible sound files from numerical (simulated, measured or synthesized) data”. The technique is used to model the sound signals and acoustics of the real environment. Virtual reality (VR) has been used in several studies [2] [3] in the past as a playback medium for accurately rendering the sonic and visual environments. Recently, it has been widely used for auralization in soundscape studies. There are two main advantages of using Virtual reality in soundscape studies. The first main advantage is high ecological validity of different soundscapes which gives the perfect illusion of “being there” for the test subject. The second major advantage is that immersive reproduction by VR enables the evaluation of soundscapes in controlled laboratory conditions. This allows soundscape researchers to isolate the different stimuli to understand the independent contributions to the perception of the entire soundscape.

Often in soundscape studies, reproduction methods used are focused on rendering the visual stimuli accurately and rely on traditional stereo or surround sound reproduction techniques for audio. The limitations of this method are the limited area of the sweet spot and expensive speaker setups. However, software tools for spatial audio such as, spatial workstation [4], and soundscape renderer [5], allow ambisonic recordings to be converted for different spatial audio playback formats like binaural and WFS. Spatial audio is advantageous over stereo and surround sound formats since it gives a more immersive feel to the user. Binaural reproduction over headphones have recently become popular

among soundscape researchers since it does not require expensive speaker setups. Binaural playback using commercial headphones is sufficient for studies which require placement of sound sources in the horizontal plane. However, there are a few challenges in using binaural playback for higher spatial fidelity, such as front back confusions and in-head localization, which occur due to limitation of speaker drivers for commonly used consumer headphones [6]. A possible solution for these challenges is to use 3D audio headphones for playback [7].

This paper details a method of performing a omnidirectional video recording of a park and playback of the spatial audio recorded using a sound field microphone. The video is stitched together using commercial software [8], and the audio is rendered using the reaper software along with spatial audio workstation provided by facebook [4]. To compare and contrast the masking effect between VR and AR, a virtual sound source in the form of a virtual speaker is simulated for testing the masking effects using augmented reality (AR) device (Microsoft Hololens [9]) and how it compares with the VR reproduction for masking in soundscape.

## 2. Visual and audio recording setup

### 2.1 Omnidirectional visual and ambisonic audio recording

#### 2.1.1 Equipment

The recording equipment for omnidirectional recording consists of six YI 4K action cameras, each recording in 2.7K resolution. The rig for mounting these cameras were custom designed and fabricated with 3D printing. On top of the rig sits the ambisonic microphone (Core Sound Tetramic). The microphone allows ambisonic recording of soundscape. The design of the omnidirectional video recording setup is shown in figure 1.



Figure 1: Omnidirectional video recording setup with 4K cameras and Ambisonic microphone

#### 2.1.2 Methodology

The methodology of recording is to use the device shown in Figure 1 for omnidirectional video and ambisonic audio recording in an urban park or any other place of interest for soundscape recording. In addition to the equipment in Figure 1, a calibrated class 1 microphone (GRAS 40PH) was used

along with a 24-bit analog-to-digital converter (NI 9234) system, controlled with a laptop running NI LabVIEW for recording the actual SPL level and raw audio data for reference.

The setup can be used for outdoor recordings both for noisy locations like the rooftop gardens or an urban park and for recording positive natural sounds for a long duration. Moreover, the data recorded by the calibrated microphone gives the accurate SPL and can be used later for post processing tasks, for example, computation of psychoacoustic parameters.

### 3. Audio and Visual rendering for VR

The captured audio and 360° video must be post processed and rendered for playback on the VR device. Thus, the process of audio rendering and playback for VR can be summarized in three steps as shown in Figure 3. The visual recordings from the 6 YI cameras shown in Figures 1 and 2 are stitched together using the Kolor Autopano Video Pro 2.5 software. The software also helps in synchronizing the audio along with the video recorded by the cameras. The spatial audio rendering is done using the reaper software [8] along with the facebook spatial workstation plugin[4]. The software and plugin allow the position of virtual sound sources to be adjusted at different spatial locations for the playback. Since the recording is made with the help of an ambisonic microphone, we can downmix the sound to binaural and use the spatial locations using the default HRTF functions in spatial audio workstation software. The additional benefit of using this approach is that we are able to add multiple sound sources at different spatial locations. This allows the control over spatial positions of both noise source and positive sound masker. The efficacy of using different types of positive sound for masking traffic noise was studied. Some of the combinations taken into consideration were effect of masking using different types of bird and water sounds along with traffic noise. The detailed subjective survey and it's results will be elucidated in a separate study.

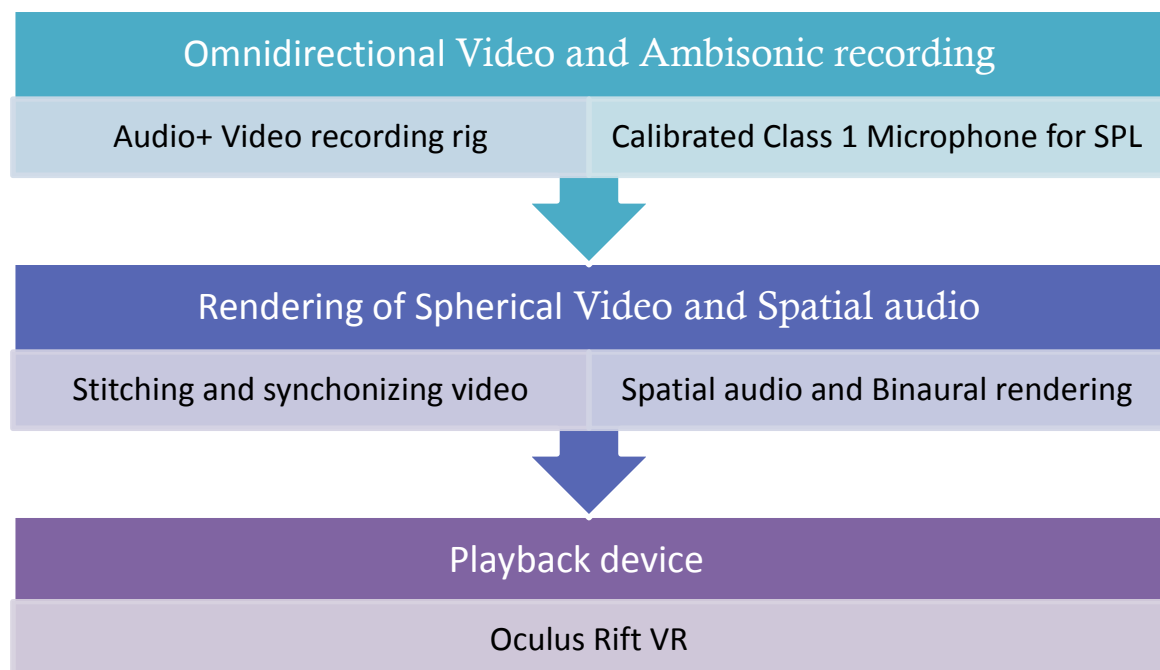


Figure 3: The complete reproduction process for omnidirectional video and spatial audio reproduction in VR

The playback device was Oculus rift VR for the rendered visual and audio. The playback device was chosen among the different available options since it has the highest resolution among all the consumer devices available currently.

#### 4. Development of soundscape evaluation tool using AR

Microsoft HoloLens is the only untethered augmented reality device (AR) that is currently available in the consumer market. It has side emitters for 3D audio reproduction and a HMD display with three processors out of which, one exclusively handles the processing for holograms.

Augmented reality offers a unique opportunity for in-situ evaluation of sound masking in soundscape. The device in this study was used for generating a virtual source modelled visually using a speaker. The application had several important features for studying the effects of spatial positioning of masker on the subsequent effect of masking traffic noise. The device used is shown in Figure 4 along with the open back headset. These open back headphones allow the traffic sounds simulated using a physical speaker to enter the ear canal together with the masker sounds generated using the HoloLens.

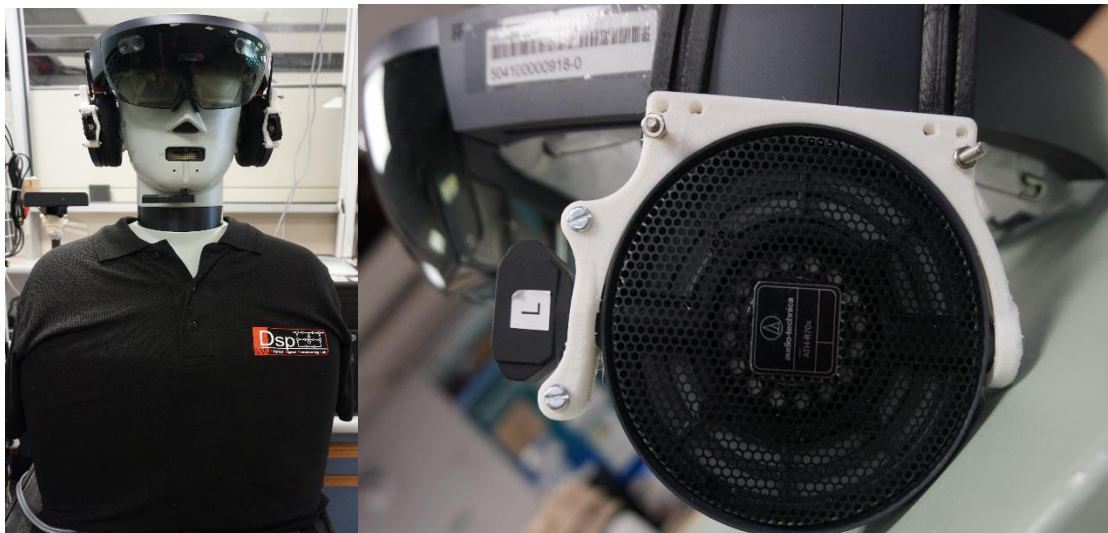


Figure 4: Microsoft HoloLens together with ATH-R70x open back headphones

The application for the virtual speaker allows the user to select and move the holograms in space. This enables the user to test the efficacy of spatial position of the masker with respect to a static noise source. A pre-amplifier device used along with the setup also allows user to adjust and monitor the gain for the masking track.

#### 5. Conclusion

In the present study, construction of a recording device for 360° video and ambisonic audio recording was described. The method of rendering audio and video using different tools is shown. Different masking tracks of recorded bird and water sounds were selected and their efficacy in masking traffic noise is evaluated using VR device. Spatial masking using a virtual speaker is studied through construction of an audio headset device along with Microsoft HoloLens AR device. This enabled the evaluation of effect of spatial effect of the masker for a static noise source simulated by a physical speaker. The user could control the gain through the pre amplifier and use hand gestures to move the

hologram of the sound object at any point in space. The devices and tools described in this paper can be used in extensive laboratory and in-situ studies for soundscape.

## Acknowledgement

This material is based on research/work supported by the Singapore Ministry of National Development and National Research Foundation under L2 NIC Award No.: L2NICCFP2-2015-5

## 6. References

- 1 M. Vorländer, *Auralization*. (2008).
- 2 T. Lentz, D. Schröder, M. Vorländer, and I. Assenmach, Virtual reality system with integrated sound field simulation and reproduction, *Eurasip Journal on Advances in Signal Processing*, **2007** (1), 187, (2007).
- 3 R. Savioja, L., Huopaniemi, J., Lokki, T., & Väänänen, Creating Interactive Virtual Acoustic Environments, *J. Audio Eng. Soc.*, **47** (9), 675–705, (1999).
- 4 Spatial workstation. [Online]. Available: <https://facebook360.fb.com/spatial-workstation/>.
- 5 K. Bredies, N. A. Mann, J. Ahrens, M. Geier, S. Spors, and M. Nischt, The multi-touch SoundScape renderer, *Proceedings of the working conference on Advanced visual interfaces*, 466–469, (2008).
- 6 K. Sunder, J. He, E. L. Tan, and W. S. Gan, Natural sound rendering for headphones: Integration of signal processing techniques, *IEEE Signal Processing Magazine*, **32** (2), 100–113, (2015).
- 7 K. Sunder, E. L. Tan, and W. S. Gan, Individualization of binaural synthesis using frontal projection headphones, *AES: Journal of the Audio Engineering Society*, **61** (12), 989–1000, (2013).
- 8 Autopano Video software. [Online]. Available: <http://www.kolor.com/2016/05/25/video-stitching-software-autopano-video-2-5-alpha-1/>.
- 9 Hololens. [Online]. Available: <https://www.microsoft.com/microsoft-hololens/en-us>.