Audio Engineering Society

Conference Paper

Presented at the International Conference on Audio for Virtual
and Augmented Reality, 2022 August 15–17, Redmond, WA, USA

VR Test Platform for Directionality in Hearing Aids and Headsets

Jesper Udesen

GN Audio, Lautrupbjerg 7, 2750 Ballerup, Denmark

Correspondence should be addressed to Jesper Udesen (judesen@jabra.com)

ABSTRACT

This paper describes how Virtual Reality (VR) is used to test the directionality algorithms in headsets and hearing aids. The headset directionality algorithm under test is based on anechoic chamber measurements of microphone impulse responses from a physical headset prototype, with 8 MEMS microphones. The algorithm is imported into Unity3D using the Steam Audio plugin. Audio and video are recorded in different realistic environments with the 4th order ambisonic Eigenmike and the 360-degree Garmin Virb camera. Recordings are imported into Unity3D and audio is played back through headphones using a virtual speaker array. Finally, the combined system is evaluated and tested in VR on human participants.

1 Introduction

Hearing aids and headsets (in the following called hearing devices) have traditionally been tested in laboratory conditions and field tests on real users. The laboratory has the benefit of being a controlled environment where the hearing device performance can be carefully investigated e.g., on a dummy head in an anechoic chamber. However, the controlled environment comes at a price. The laboratory tests are often far from the reality a human user will experience with the hearing device on his ears and the ecological validity is low. Therefore, the laboratory tests are supplemented with field tests where real humans test the hearing devices in everyday usage. Here, there is a high degree of ecological validity, but the test data will be subject to noise sources that are difficult to control and quantify. Hence the field test data are challenging to reproduce. Furthermore, the field tests are often conducted a long time (sometimes years) after the first laboratory tests. This makes it time-consuming and expensive to optimize a given hearing device if several cycles of laboratory tests and field tests are needed.

It has been proposed to close the gap between unrealistic laboratory tests and uncontrolled field tests using an advanced speaker array system combined with a VR headset (e.g.,[1][2]). The playback audio can be simulated or prerecorded using a higher-order ambisonic (HOA) microphone. Visual playback in the VR headset must match the audio and can be prerecorded using a 360-degree video camera. This setup requires the test subject to sit in the center of the speaker array with the hearing device on his ears. However, such speaker array systems are often costly and they should ideally be placed in an anechoic chamber. Also, the human test subject must wear a real physical hearing device which has to be designed and built; a process that can take several months and involves hours of laboratory testing.

This paper investigates the possibility to circumvent the expensive speaker array and the physical construction of the hearing device by using a *virtual* hearing device and a *virtual* speaker array

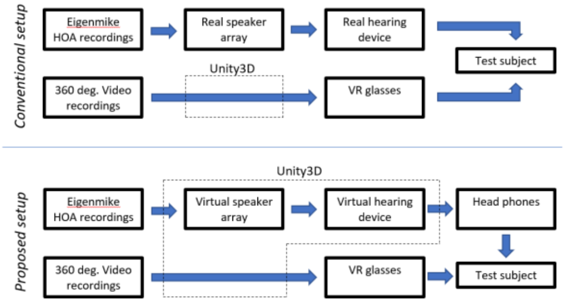

Fig. 1. The proposed processing flow

implemented in Unity3D and let the audio be played back through ordinary headphones. The processing flow of the conventional setup and the proposed setup are both illustrated in Figure 1. The proposed setup could allow researchers to do ultra-fast prototyping of new ideas and concepts in a controlled VR environment. Furthermore, the hardware in such a setup will be limited to a standard gaming laptop, a VR headset, and a pair of headphones.

As a test example, the virtual hearing device will emulate a directionality algorithm in a headset prototype with 16 microphones. The directionality algorithm attenuates sound coming from the rear direction while preserving sound from the front of the user. Directionality algorithms are standard in modern hearing aids[3][4].

This paper is organized as follows: In Section 2 the virtual hearing device is described. In Section 3 the HOA recordings, the 360 degrees video recordings, as well as the virtual speaker array, are described. Finally, in Section 4 the combined system is evaluated in Unity3D and tested on 10 test subjects.

2 The virtual hearing device

The virtual hearing device is designed in four steps: 1) Build a physical prototype with microphones, 2) Anechoic room measurements on the prototype, 3) Directivity optimization of measured data, 4) calibration of data to meet the *flat-insertion-gain* criteria, 5) convert the impulse response data to SOFA file format [5] and import into the Steam Audio Plugin in Unity3D. Each of these steps will be described in the following.

The virtual hearing device is based on measurements with a physical prototype with 8 MEMS MM20 microphones from Knowles Electronics attached to the right earcup of a Jabra Evolve 80 headset. The headset was only used as a form factor for the microphones and the headset loudspeaker was not used. The prototype (Figure 2) was placed on a HATS (Head And Torso Simulator) which was placed on a B&K 9640 turntable in an anechoic room. Eight homemade loudspeakers fitted with 2.5-inch drivers were mounted in an arc at a distance of 1.5 m from the center of the HATS head at elevation angles {90°, 67.5°, 45°, 22.5°,0°,-22.5°,-45°,-67.5°}. The impulse responses from each speaker to each of the MEMS microphones were measured at a sampling frequency of 48 kHz with a horizontal resolution of 5° controlled by the turntable. The excitation signal was a code length 11 maximum length sequence (MLS) [6] signal with a duration of 5 seconds for each measured impulse response.

The measured impulse responses were calibrated to remove any effects of a non-flat speaker response. This was done by convolving each impulse response with the inverse of the speaker responses. The speaker

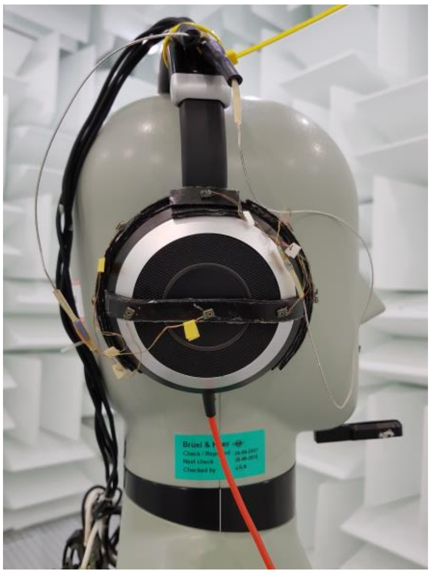

Fig. 2. The hearing device prototype is seen on a HATS. The microphones are connected to external hardware for measurements.

responses were measured in the center of the arc (when HATS was removed) with a ½-inch B&K reference microphone.

The recorded data consists of 72x8x7+1=4033 impulse responses (72 horizontal angles, 8 microphones, and 7 speakers plus one impulse response for the speaker at 90° elevation). These data were extended to the left earcup of the Evolve 80 headset assuming symmetry of the HATS head. This increased the total number of microphones to 16 and the total number of impulse responses to 4033x2=8066.

The hearing device uses all 16 microphones to filter the incoming sound, add all the filtered signals together and output a mono signal which is played back to both ears, i.e., it is a classical filter-and-sum beamformer. The beamformed response for an input sound originating from a speaker at an angle θ_i playing a Dirac delta function can be written in the time domain as:

$$\bar{z}^{\theta_i}(n) = \sum_{m=1}^{M} \bar{h}_{mic_m}(n) * \bar{x}_{mic_m}^{\theta_i}(n) \qquad (1)$$

where $\bar{z}^{\theta_i}$ is the filtered output, M is the number of microphones (in this case 16), $\bar{h}_{mic_m}(n)$ is the beamforming filter applied to microphone m and $\bar{x}_{mic_m}^{\theta_i}$ is the measured impulse response for microphone m and (*) is the convolution operator.

The beamforming filters $\bar{h}_{mic_m}$ are found by solving the following least-square optimization problem:

$$\begin{aligned}&\bar{h}_{mic_1}(n), \bar{h}_{mic_2}(n), \ldots, \bar{h}_{mic_M}(n)\\ &= \arg\min \sum_{n=1}^{N} \sum_{i=1}^{I} \left(\bar{y}_{Des}^{\theta_i}(n) - \bar{z}^{\theta_i}(n)\right)^2\end{aligned} \qquad (2)$$

where $\bar{y}_{Des}^{\theta_i}(n)$ is the desired response at a given angle. For this prototype, the desired response was set to be the sum of all measured impulse responses weighted with a Hanning function with a maximum at (0°,0°) and a width of 60° in both azimuth and elevation. The solution to Equation (2) can be found from [7]:

$$\begin{bmatrix} \bar{h}_{mic_1}^T \\ \bar{h}_{mic_2}^T \\ \vdots \\ \bar{h}_{mic_M}^T \end{bmatrix} = \begin{bmatrix} \bar{\bar{R}}_{mic_1}^{\theta_1} & \bar{\bar{R}}_{mic_2}^{\theta_1} & \cdots & \bar{\bar{R}}_{mic_M}^{\theta_1} \\ \bar{\bar{R}}_{mic_1}^{\theta_2} & & \cdots & \cdot \\ \vdots & & & \\ \bar{\bar{R}}_{mic_1}^{\theta_I} & & \cdots & \bar{\bar{R}}_{mic_M}^{\theta_I} \end{bmatrix}^{-1} \begin{bmatrix} \bar{y}_{Des}^{\theta_1} \\ \bar{y}_{Des}^{\theta_2} \\ \vdots \\ \bar{y}_{Des}^{\theta_I} \end{bmatrix} \qquad (3)$$

where $\bar{\bar{R}}_{mic_m}^{\theta_i}$ is the convolution matrix for $\bar{x}_{mic_m}^{\theta_i}$.

The estimated spatial responses $\bar{z}^{\theta_i}$ of the hearing device can be found by substituting the beamforming filters $\bar{h}_{mic_m}$ from Equation (3) into Equation (2). The corresponding responses for elevation angle 0° are plotted at the top of Figure 3. Most of the acoustic energy is focused in a zone around the 0° azimuth angle (as defined by the desired response). A characteristic spatial aliasing pattern can be seen. The pattern is a result of constructive and destructive interference due to the distance between the two earcups. This effect is a well-known physical limitation of multi-microphone arrays [8]. The bottom part of Figure 3 shows the Directivity Index (DI) [8] of the hearing device. The DI is approximately 10 dB higher than the corresponding DI of the open ear (the 711 Coupler response) of HATS.

In a physical prototype, it will not be possible to realize such a high DI at low frequencies due to the *white-noise-gain* problem where internal microphone noise is amplified [7][8]. However, the virtual hearing device is not limited by such constraints due to the high SNR on the microphone impulse responses obtained by the MLS decoding. If one were to investigate the effect of microphone noise using the virtual hearing device, the true microphone noise should be added to the individual microphone impulse responses $\bar{x}_{mic_m}^{\theta_i}$.

The estimated spatial responses $\bar{z}^{\theta_i}$ are optimized to the desired target function $\bar{y}_{Des}^{\theta_i}$ but this does not guarantee that the hearing device sounds “natural”. To achieve this, a final processing step is needed where $\bar{z}^{\theta_i}$ is matched to the open ear response of the HATS for a sound source at the target direction (0° azimuth). This is the so-called *flat-insertion-gain* calibration [3][4].

The calibrated spatial responses $\bar{z}^{\theta_i}$ characterizes the spatial sensitivity of the hearing device in the same way a head-related-impulse-response (HRIR)

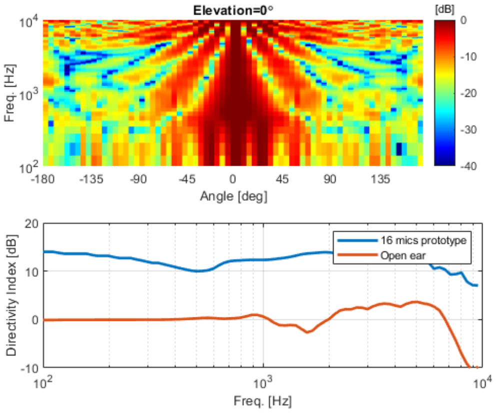

Fig. 3. Top: The hearing device response in the horizontal plane. Bottom: The corresponding directivity index compared to the open ear directive index of the HATS.

characterizes the spatial sensitivity of the human ear. We can therefore replace the HRIR database in Unity3D with the spatial responses $\bar{z}^{\theta_i}$ which are 512 taps long. In practice, this is done by saving all $\bar{z}^{\theta_i}$ data in SOFA format [5] and importing data into the Steam Audio plugin in Unity3D.

3 The HOA and 360-degree video recordings and the virtual speaker array

A 4th order ambisonic microphone from mh-acoustics (the Eigenmike) was used to record two-minute clips in 14 different environments (including canteen, traffic, meeting, and cocktail party). Each clip also included 360-degree video recordings at 30 fps with the Garmin Virb 360 camera (5.7K resolution). The Garmin Virb camera was attached to the top of the windscreen of the Eigenmike which reduced the distance to 10 cm between the camera and the microphone.

The 4th order ambisonic recordings were decoded to a speaker array using a Matlab ambisonic decoder with *Sampling-Ambisonic-Decoding* (SAD) [9]. The speaker array geometry matched a real array of 39 speakers (Figure 4) located in a semi-anechoic room at GN, Ballerup, Denmark. Array radius was 1.5 m and speaker positions were defined by four horizontal “rings” of equidistant speakers: elevation -30°, 10 speakers; elevation 0°, 12 speakers; elevation 30°, 10 speakers; elevation 60°, 6 speakers. Finally, at an elevation angle of 90°, there was one speaker mounted on the ceiling. For all the four “rings” there was a speaker at 0° azimuth. The physical speaker array was replicated in Unity3D with each speaker being a Unity audio source. The playback signal for each Unity speaker matched the playback signal of the real physical speaker array.

4 Tests in Unity3D

The virtual speaker array and the virtual hearing device were tested in Unity3D in three steps: 1) a single audio source response, 2) all 39 speakers playing the same signal, 3) perceptual tests with the HOA recordings and 360-degree video recordings to check for dynamic artifacts. Each of these steps will be described in the following.

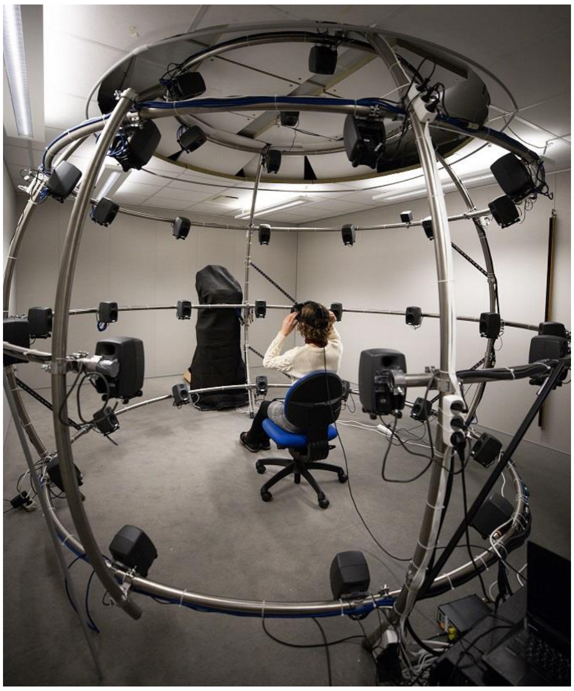

Fig. 4. The speaker array replicated in Unity3D

The spatial responses $\bar{z}^{\theta_i}$ derived in Section 2 were loaded into the Steam Audio plug-in in Unity3D, and a single Unity audio source was playing a code length 12 MLS signal of 5 seconds duration. The audio source was placed at an elevation angle of 0° relative to the audio listener and rotated around the audio listener in steps of 5° azimuth. The resulting signals were recorded and decoded and the corresponding data can be seen at the top of Figure 5 which can be directly compared to Figure 3. The bottom of Figure 5 shows the (0°,0°) response from Unity3D together with the spatial response $\bar{z}^{\theta_i}$ for (0°,0°). The absolute difference between the two curves in the bottom of Figure 5 has a standard deviation of 1.3 dB when measured in third-octave bands between 20 Hz and 20 kHz. The deviation between the two curves is small (<1 dB) at low frequencies and gets larger above 10 kHz.

When using Unity3D for audio playback of ambisonic signals, it is important that all virtual speakers in the array are excited at the same time and there are no time misalignments. To check if this was the case, all 39 speakers in the speaker array defined in Section 3 were playing the same code length 12 MLS signal of duration 5 seconds. The HRIR database in the Steam Audio plugin was changed to be a Dirac delta function for all angles and the audio signal was recorded by an audio listener in the center of the array. The recorded signal was decoded, and it was found that the signal was a replica of the input signal but with an amplitude 39 times higher due to constructive interference of each audio source. This shows that Unity3D does not introduce time misalignments on the audio signals.

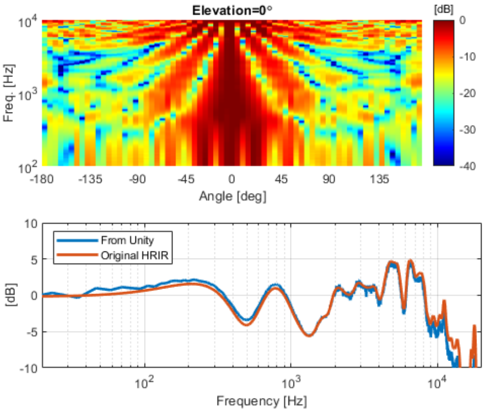

Fig. 5. Top: The hearing device response in Unity3D to a single audio source in the horizontal plane (shown without flat-insertion-gain for easy comparison with Figure 3). Bottom: The hearing device response in Unity3D to a single source at (0°,0°) vs the spatial response $\bar{z}^{\theta_i}$ for (0°,0°)

Dynamic artifacts may occur if the spatial sampling of the HRIR data is too coarse [6]. In the present study, the elevation angle between HRIR data is 22.5° (determined by the “arc” of speakers in the anechoic room measurements). It could therefore be expected that artifacts would be present when the test subjects rotated their heads during playback. To check for dynamic artifacts 10 test subjects tested the full setup using an Oculus Rift S VR headset and a pair of Sennheiser HD 650 for audio playback. The test subjects could switch between the 14 recorded environments (both visuals and audio) and change between the virtual hearing device with 16 microphones and a standard open ear HRIR database for reference. It was reported by the test subjects that there were no audible artifacts due to head rotations. Furthermore, it was reported that the hearing device with the 16 microphones had a significant “beam” effect which was to be expected due to the DI improvement of ~10 dB relative to the open ear (Figure 3).

Some of the recorded environments included rooms with significant room reverberation times. Here it was reported by the test subjects that the hearing device attenuated room reflections more than the open ear mode which should also be expected for a high directivity beamformer.

The hearing device takes the 16 microphones as input and outputs a single mono signal without any binaural spatial cues (ILD, ITD). This lack of spatial cues was noted by most test subjects as a lack of spatial unmasking.

5 Conclusion

This study tested if Unity3D and the Steam audio plugin can be used to simulate the effect of a directionality algorithm. It was found that the audio output of Unity3D for a single stationary sound source matched the expected output except for a small deviation (1.3 dB) for which the author has no explanation. The dynamic performance of the system was tested on 10 test subjects and user feedback indicates that the system works as intended.

The virtual hearing device and the virtual speaker array do not have imperfections that can be found in a real speaker array and a physical hearing device. In a physical speaker array for ambisonic playback, it can be difficult for the test subject to stay within the sweet spot of the system due to head movements. In the virtual setup, the test subject is fixed to the center position and thereby also to the center of the sweet spot.
In the physical speaker array, the test subject will be wearing a VR headset which will change the sound field due to diffraction and reflection of sound. It has previously been shown that this effect is small [1] but with the virtual speaker array, such problems can be avoided completely.

The virtual hearing device and the virtual speaker array allow the researcher to do fast prototyping without having to deal with an expensive speaker array and a physical real-time prototype hearing device. In the present study, this was exemplified with a directionality algorithm but also compression and noise reduction algorithms could be implemented, thereby emulating the core components of a hearing aid. Such a setup can be combined with classical user tests like *paired-comparison* or *MUSHRA* running in VR. This would allow the researcher to perform quantitative user tests at a very early stage in the development process.

References

[1] A. Ahrens, K. D. Lund, M. Marschall, and T. Dau, “Sounds source localization with varying amount of visual information in virtual reality”, *PLoS ONE,* 14(3), (2019)

[2] T. Huisman, A. Ahrens and E. MacDonald, “Ambisonics Sound Source Localization With Varying Amount of Visual Information in Virtual Reality”, *Frontiers in Virtual Reality*, vol. 2, (2021)

[3] H. Dillon, “Hearing Aids”, Thieme Medical Publishers Inc, (2012)

[4] J. M. Kates, “Digital Hearing aids”, Plural Publishing Inc, (2008)

[5] https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics)

[6] B. Xie, “Head Related Transfer Function and Virtual Auditory Display”, J Ross Publishing, (2013)

[7] I. J. Tashev, “Sound Capture and Processing”, Wiley, (2009)

[8] B. Rafaely, “Fundamentals of Spherical Array Processing”, Springer, (2015)

[9] A. Politis, “Microphone array processing for parametric spatial audio techniques”, Doctoral Dissertation, Department of Signal Processing and Acoustics, Aalto University, Finland (2016)