HMD-based Virtual Environments for Localization Experiments

Tim Lübeck[1], Johannes M. Arend[1,2], Christoph Pörschmann[1]
[1] Institute of Communications Engineering, TH Köln, D-50679 Cologne, Germany
[2] Audio Communication Group, TU Berlin, D-10587 Berlin, Germany
Email: tim.luebeck@th-koeln.de

Introduction

Virtual acoustic environments (VAEs) have already become an accepted research tool for performing psychoacoustic experiments. Mostly, scientific experiments were held under slightly varying conditions. Researchers from different laboratories use their own dedicated hard- and software, utilize different techniques for presenting the VAE, or develop own ways of reporting the participants' judgment. These specific, and mostly unknown conditions, complicate the performance of transparent and reproducible experiments. The technical quality of virtual reality (VR) systems, e.g. Oculus Rift or HTC Vive, and especially head-mounted displays (HMDs) has made great progress over the last years. These standardized commercial devices are less expensive than established tracking systems and appear to be worthwhile tools to setup unified scientific experiments.

Applying HMD-based virtual environments (VEs) for psychoacoustic experiments has a number of additional advantages. VR systems provide an accurate and trustworthy full-spherical tracking system, extendable with additional tracking devices. Several open source development tools, as for example the OpenVR SDK, Steam Audio, or Unity3D make it easy to implement VEs. Finally, the presentation of a visual environment can simply be integrated.

For localization experiments, accurate pointing to the perceived auditory event is a crucial issue. Thus, many different pointing methods have been developed in various studies on human sound source localization, (see e.g. [1], [2], [3]). Either for loudspeaker or headphone-based experiments, two different egocentric methods have been proven to be most accurate and applicable: the finger and the head pointing method. If the finger pointing method is applied, the participants have to point with their fingertips towards the perceived sound source. In experiments applying head pointing, the participants have to direct the head towards the position of the auditory event. However, in order to perform HMD-based localization experiments, an appropriate VR pointing method has to be found. Inspired by the egocentric techniques, we developed a so-called laser pointing method. Here, the participants hold the controllers in their hands, and at the top of one of them a laser beam appears inside the virtual environment. With this beam, the participants can control a marker, for example, a white sphere, and place it at the perceived position of the auditory event. This enables free pointing in any direction and distance.

This work was funded by the German Federal Ministry of Education and Research (BMBF 03FH014IX5-NarDasS)

In this work, we present a first approach for performing HMD-based listening experiments. To examine the applicability of HMDs for listening tests and to evaluate our new pointing method, we conducted a localization experiment, comparing the finger, head and laser pointing method. Using the 3D game engine Unity3D and the Oculus Rift HMD and controller bundle, we implemented a headphone based experiment involving the three pointing methods. The experimental design is based on the study of Bahu et al. [1], who compared a head, finger and proximal pointing in a loudspeaker based localization experiment. Three groups of participants, each assigned to one of the pointing methods, participated in the experiment. All participants had to rate the same target sound source positions with their respective pointing method. The results of this experiment indicate whether HMDs are suitable tools for psychoacoustic experiments and if our proposed pointing method achieves similar results as established methods. Additionally, we discuss various new possibilities HMD-based VEs offer for psychoacoustic experiments.

Method

Participants

Ten females and 44 males aged between 19 to 30 years (M = 22.94 years, SD = 2.43 years) participated in the experiment. Most of them were students in media technology at TH Köln with minor experience in binaural synthesis or HMD-based VR systems. They were divided into three evenly distributed pointing-method groups of 18 participants. Thus, each group of participants had to rate the same conditions but with a different pointing method.

Setup and Stimuli

The experiment took place in an acoustically damped laboratory room at TH Köln with a background noise level of about 32 dB(A). The experimental conditions were set up in MATLAB. This involves the definition of the sound source positions or the randomization of the trials. This information is transferred to Unity3D to execute the experiment. For auralization, we used the SteamAudio plugin, which enables dynamic binaural synthesis with arbitrary HRTFs (Head-Related Transfer Functions) provided in the SOFA format [4]. The applied HRTFs are based on spherical measurements of a Neumann KU100 dummy head made on a Lebedev grid

with 2702 nodes [5]. As this set is stored in the spherical harmonics domain in the form of SH-coefficients, HRTF sets on an arbitrary sampling grid can be generated by means of spherical harmonics interpolation. We generated an HRTF set of discrete source positions with a resolution of 2° in the horizontal plane and a resolution of 5° in the median plane. This seemed entirely sufficient for the purpose of this experiment. Furthermore, we disabled any reflections, distance properties, or room acoustics in the SteamAudio settings, and thus presented only pure positional rendering based on convolution with the HRTFs. For interpolation between any HRTF grid point, SteamAudio provides a next neighbor interpolation which we applied in this case. The stimuli were played back via Sennheiser HD 600 headphones and an RME Babyface as A/D converter with an equivalent sound level of about 60 dB(A). The headphones were compensated by inverse filtering according to Erbes et al. [6]. For tracking and presentation of the visual environment, we used the Oculus Rift VR kit version *CV*1, and SteamVR which is based on the OpenVR SDK.
Similar to the study from Bahu et al. [1] we presented the virtual sound sources at positions according to Table 1. The anechoic test-signal consisted of three 100 ms white Gaussian noise burst (including 10 ms cosine-squared onset/offset ramps) and pauses of 30 ms.

Table 1: Presented sound source positions. Similar to Bahu et al. [1], we tested in total 24 positions. All elevation (El.) and azimuth (Az.) angles specified in degree (°).

El.	Az.								
-5	-160	-120	-80	-40	0	40	80	120	160
25	-20	-60	-100	-140	180	140	100	60	20
60	-144		-72		0		72		144
90					0				

Procedure

The participants sat on a swivel chair, while wearing the HMD, the controllers, and the HD 600 headphones. As visual environment we presented a dark shoebox room with the dimensions 100 m×100 m×100 m. The floor, the ceiling, and the walls were covered with a blue grid, as depicted in Figure 1. At the beginning of each trial, the participants were instructed to orient to the front, which was checked by the orientation of the HMD. If this was the case, the stimulus was played back. Although dynamic binaural synthesis was used, the participants were asked to omit head movements during the playback, which was also monitored. After the playback, the participants were free to rotate with the chair. To report the perceived sound source position, the participants used their respective pointing method, and confirmed the judgment with a button click. After turning back to the initial front direction the next stimulus was presented.
To get used to the VR environment, the handling of the controllers, and the experimental procedure, all participants had to perform a short training with 10 trials beforehand. After that, the actual test started with a total of 192 trials (24 conditions with 8 repetitions each). After half of the trials, the participants could make a short break.

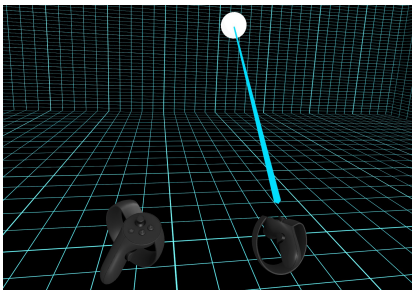

Figure 1: Screenshot of the HMD view, which displays both controllers. At the right controller a laser beam with the attached marker approaches. Furthermore, the dark shoebox environment with the rectangular grid is illustrated.

Data Analysis

In previous studies, various measures to describe the localization accuracy, have been established. We determined the accuracy by the unsigned angular deviation, calculated by the dot product between target and judgment direction. Furthermore, we considered the horizontal and vertical error separately in the analysis. These errors were calculated as the unsigned azimuth and elevation angular deviation.
A known effect of missing dynamic cues while presenting binaural synthesis is the interchanging of sources at the front and back, or top and down. These positions do not lead to significant interaural time differences, and yield to so-called confused judgments [7]. According to Wightman and Kistler [8], we detected the confused judgments as follows: deviations of the target and judged direction greater than the deviation of the judged and the target direction mirrored at the frontal plane was indicated as confused judgment. Just like Bahu et al. [1], we corrected these judgments by mirroring them at the frontal plane, before applying further data analysis.
The statistical analysis is based on the mean angular deviation of all repetitions, averaged over all participants per pointing group. We excluded the judgments for 90° elevation. A Lillifors test for normality failed to reject the null hypothesis for 4 of 18 conditions at a significance level of 0.05. However, parametric tests as ANOVA are generally robust to violations of normal distribution assumption. We thus performed a three-way mixed ANOVA with Greenhouse-Geisser corrections for violations of the assumption of sphericity according to the 3 × 3 × 2 mixed factorial design. The between-subjects factor was pointing method (finger pointing, head pointing, laser pointing) and the within-subjects factors were elevation (-5°, 25°, and 60°) and target hemisphere (front, back). For further post-hoc analysis, various independent samples t-tests between each subject group were applied. We additionally performed two more mixed ANOVAs using the same between and within-subject factors as described above, but for the horizontal and vertical error as

accuracy measure. The Lillifors test failed to reject for 3 of 18 conditions for the elevation error, but never failed for the azimuth error.

Results

When considering target positions in the frontal and rear hemisphere separately, we obtained an average frontal confused judgment rate of 61.40% (Head Pointing (HP): 63.66%, Finger Pointing (FP): 57.45%, Laser Pointing (LP): 63.06%), an average confused judgment rate of 8.07% for the back (HP: 5.93%, FP: 12.55%, LP: 5.74%), and an overall average confused judgment rate of 34.73% (LP: 34.4%, FP: 35%, HP: 34.8%).
The ANOVA for the average angular deviation yielded a significant influence of the pointing method ($F(2,51) = 4.53$, $p = .015$, $\eta_p^2 = .151$). Subsequent independent samples t-tests between the averaged angular deviations of each pointing group showed that the FP method has the strongest influence (between FP and LP: $t(862) = 5.851$, $p < .001$, $d = 0.398$; between FP and HP: $t(862) = 3.763$, $p < .001$, $d = 0.256$; between HP and LP: $t(862) = 2.00$, $p = .046$, $d = 0.136$). Figure 2 shows the average angular error of all subjects for each pointing method, including the 95% between-subject confidence intervals. The group of participants using the FP method achieved in average an error of 40.1° , and performed around 7° more imprecise than the LP method group with an average error of 32.01°. Participants using the HP method achieved an average deviation of 35.2°.

Furthermore, the ANOVA revealed an interaction effect between the pointing method and the elevation angle ($F(4,102) = 3.223$, $p = .037$, $\eta_p^2 = 0.478$, $\epsilon = 0.614$). We thus observed a slight influence of the pointing method dependent on the elevation angles. Figure 3 presents the average deviation with respect to the elevation error for the front and back hemisphere. It shows that independent of pointing method, high elevation angles yielded worse results. The HP and FP groups performed rather similar, whereas the LP method group achieved around 10° more precise estimations.

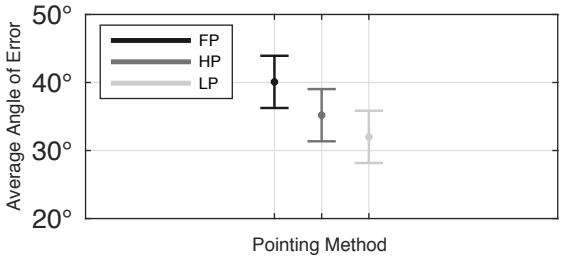

Figure 2: The angle of error, averaged over all conditions and all participants, for each pointing method. The error bars mark the 95% between-subject confidence intervals. All methods lead to a mean accuracy between 30°and 40°. Participants using the FP method achieve in average around 7° worse results. As overal average deviation, we obtain 35.2° for participants using the HP method, 40.1° for participants using the FP method, and 32.01° for the LP method group.

Moreover, we consider the vertical and horizontal error separately. The ANOVA for the vertical error showed a moderate main effect for the pointing method ($F(2,51) = 3.263$, $p = .046$, $\eta_p^2 = 0.113$), as well as an interaction effect between pointing method and hemisphere ($F(4,102) = 4.692$, $p = .008$, $\eta_p^2 = 0.118$, $\epsilon = .61$). Figure 4 (top) shows the corresponding mean vertical errors. The mixed ANOVA for the horizontal error yielded no significant effects. Figure 4 (bottom) shows the mean azimuth error for each elevation angle, and again for the front and back hemisphere separately. It can be seen that, unlike the elevation error, the results for front and back are completely different. For the back hemisphere, sound sources from 25 °elevation were localized the best, sources from -5° and 60° nearly the same (around 20° deviation).
Additional to the localization accuracy, we determined the mean response time per trial. For the HP method we obtained 3.22 seconds, for FP 2.98 seconds and for LP 3.03 seconds.

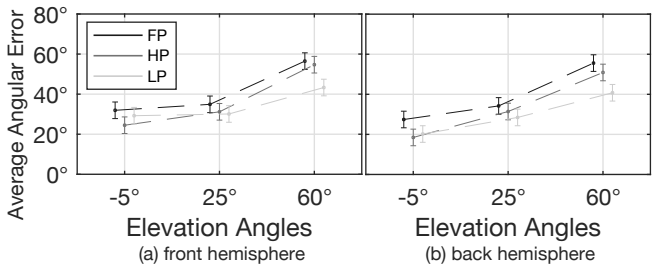

Figure 3: Average angle of error with respect to the elevation angle dependent on the pointing method for the front hemisphere (a) and the back hemisphere (b).

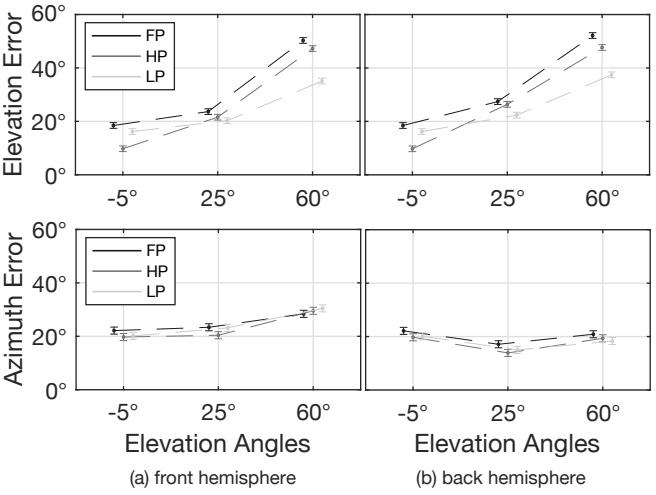

Figure 4: The elevation dependent vertical (top) and horizontal (down) angular error for front (a) a back (b) hemisphere. The elevation errors for front and back looks quite the same, whereas the azimuth error for the back hemisphere is smaller than for the front. The pointing method had no influence on the azimuth error, but there are differences in the elevation error plot. The LP method led to the best elevation accuracies. for higher elevation angles.

Discussion

In the presented study, we mainly investigated if HMD-based VR systems are an appropriate technique for performing experiments on human sound source localization.

We achieved an average angular error of 32.01° for the participant group using the LP method, 35.2° for the HP group, and 40.1° for the FP group. These error values are in the same range as the results in comparable studies using headphone-based presentations. Djelani et al. [2], who compared multiple pointing methods, indicated an average deviation of 21.26° for a finger pointing method and 20.9° using a head pointing method. Wightman and Kistler [8] reported average errors between 15.1° and 31.2°. Just like in our experiment, both studies did not permit head movements during the presentation of the stimulus. Unlike the present study, however, it should be noted that both studies used individual HRTFs. Majdak et al. [3] performed an HMD-based localization experiment and achieved horizontal errors of 15.1° to 18.6° , and vertical errors of 37.5° to 37.9°. Our results are around 4° worse for the horizontal error (24°), but about 10° better for the elevation error (27°).

Although our localization accuracy results are comparable to other studies, we observed a relatively high confused judgment rate. Markous and Middlebrooks [9] indicated an average rate of 6%, we obtained in average a rate of 34.73%. This may have been caused by using non-individual HRTFs.

Considering the influence of the pointing method, we come to a similar conclusion as Bahu et al. We indeed noticed a small influence of the pointing method, however, this is quite weak. For our experiment, the LP method group performs around 7° better than the FP group, the HP group lies in between. Bahu et al. reported an average difference of 2° between a proximal and a finger pointing. For higher elevation angles, the LP has a notable advantage over HP and FP.

In addition to the localization accuracy, some other characteristics of the pointing methods have to be taken into account. Most of the participants describe the FP method as uncomfortable, which was also discovered by Djelani et al. [2] or Majdak et al. [3]. Apart from the handling, an imprecision in determining the exact judged position can occur with the FP method. The judgment vector can be determined either by the connection of the point between the eyes and the fingertip, or the middle of the head and the middle of the hand, just to name two possibilities. Likewise, the HP method can lead to imprecise results. The participant is obligated to exactly direct the head towards the sound source, instead of just looking towards the source with the eyes. Finally, HP and FP provide no possibility for distance estimation. This is the main advantage of the LP method.

Conclusion

Overall, we showed that HMD-based VR systems are applicable for localization experiments and that similar results as in other established studies can be obtained. We developed an appropriate method for reporting participants localization estimations, which achieves slightly more accurate results as the more conventional pointing methods head and finger pointing. Nevertheless, finger and head pointing are applicable in VR too. We further demonstrated further technical opportunities of VR experiments, as presenting a visual environment or storing arbitrary tracking data. Finally, SteamAudio provides the auralization of any HRTF set in SOFA format, which simplifies the integration of established HRTF data into HMD-based environments.

Up to now, we did not examine the influence of the visual environment. In this study, we decided to present a dark shoebox room with a rectangular blue grid for orientation. Majdak et al. performed a localization experiment comparing sound source localization in complete darkness with an HMD-based presentation. They noticed a significant influence of the visual presentation. Further investigations on the interaction of auditory and visual perception, like Zalles et al. [10] or Werner et al. [11], came to similar findings. The audiovisual convergence affects the externalization, the confused judgment rate, and thus, the localization performance. We plan to conduct further studies on this topic too.

References

[1] Bahu, H., Carpentier, T., Noisternig, M., and Warusfel, O., "Comparison of different egocentric pointing methods for 3D sound localization experiments," *Acta Acustica united with Acustica*, 102(1), 2016.

[2] Djelani, T., Pörschmann, C., Sahrhage, J., and Blauert, J., "An Interactive Virtual-Environment Generator for Psychoacoustic Research II: Collection of Head-Related Impulse Responses and Evaluation of Auditory Localization," 2000.

[3] Majdak, P., Goupel, M. J., and Laback, B., "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, perception & psychophysics*, 71(3), 2009.

[4] Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., and Noisternig, M., "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions," *134th Audio Engineering Society Convention 2013*, 2013.

[5] Bernschütz, B., "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," *Fortschritte der Akustik – AIA-DAGA 2013*, 2013.

[6] Erbes, V., HagenWierstorf, Geier, M., and Spors, S., "Free database of low-frequency corrected head-related transfer functions and headphone compensation filter," *142nd Convention Audio Engineering Society*, 2017.

[7] Blauert, J., *Spatial Hearing*, Hirzel Verlag Stuttgart, Camebridge, 1997.

[8] Wightman, F. L. and Kistler, D. J., "Headphone simulation of free-field listening. II: Psychophysical validation," *The Journal of the Acoustical Society of America*, 85(2), 1989.

[9] Makous, J. C. and Middlebrooks, J. C., "Two-dimensional sound localization by human listeners," *The Journal of the Acoustical Society of America*, 87(5), 1990.

[10] Zalles, G., Genovese, A., Flanagan, P., and Roginska, A., "Evaluation of Binaural Renderers: Externalization, Front/Back and Up/Down Confusions," *AES 144th Convention*, (November), 2018.

[11] Werner, S., Klein, F., Mayenfels, T., and Brandenburg, K., "A summary on acoustic room divergence and its effect on externalization of auditory events," *2016 8th International Conference on Quality of Multimedia Experience, QoMEX 2016*, 2016.