Virtual reality based pointing method for localisation experiments in spatial audio

Vera Erbes[1], Andreas Fleck[2], Sascha Spors[1]

[1] *Institute of Communications Engineering, University of Rostock, 18119 Rostock, Germany, email: {vera.erbes, sascha.spors}@uni-rostock.de*

[2] *Ingenieurbüro Fleck, 26506 Norden, Germany, email: mail@ib-fleck.de*

Introduction

As the accurate placement of sources is an important quality in spatial sound reproduction, listening tests strive to investigate the localisation of virtual sources. The applied pointing method, which the test subjects report the perceived direction with, is crucial for meaningful results. Several studies revealed that pointing with the head in direction of the source while being assisted by visual feedback for the direction delivered the most accurate results. In particular, providing visual feedback by a head-mounted display (HMD) has been shown to be a promising method which also enables the investigation of the localisation of elevated sources. This study combines the virtual representation of sound fields by dynamic binaural synthesis via headphones with the use of an HMD for visual feedback. Stimuli consisted of point sources synthesised by head-related transfer functions (HRTFs). Comparison with the results of previous studies proves that the method provides a valid instrument for the investigation of localisation properties of spatial reproduction methods.

Localisation experiments in spatial sound reproduction

Investigation of localisation properties of spatial sound reproduction methods usually requires a variation of the loudspeaker setup, the listener position or even the listening room properties. This is difficult to realise in listening tests as small differences between conditions can only be revealed by instantaneous comparisons. Therefore, binaural synthesis is typically used to simulate different setups under test and has proven to be a transparent method for localisation studies [1].

Also critical for localisation experiments is the choice of the reporting method which has been investigated by several studies, e.g. [2, 3, 4]. Existing methods range from graphical indications on maps [5] to pointing with a hand-held device [6] or with the head [7] in direction of the auditory event. More technically advanced methods use tracking of eye movements [8] or provide visual feedback in virtual reality environments [9, 10]. Best results could be achieved with methods that combine pointing with the head supported by visual feedback of the head direction as it reduces errors induced by interaction with the locomotor system [2].

In this study, the use of dynamic binaural synthesis is combined with the reporting method of pointing with the head supported by visual feedback with an HMD, similarly to the study by Majdak et al. [9]. In [9], static binaural synthesis was used, though, where subjects had to indicate the direction of the auditory event after presentation of the stimulus. The aim of the present study is to investigate the chosen method regarding its accuracy. To this end, it is compared to results from the literature and to two previous studies [1, 11] that used exactly the same stimuli and listening test design, but different head tracking devices (Polhemus Fastrak in [1] and NaturalPoint OptiTrack in [11]) and especially a different way of providing the visual feedback. Both [1] and [11] used a laser pointer attached to the headphones on the subjects' heads projecting on a curtain. While [1] used a straight curtain, [11] could reduce still existing undershoots of reported answers for lateral directions by employing a circular curtain.

Experiment with a virtual reality based pointing method

Stimuli

Subjects had to report the direction of the auditory events for 11 source directions ϕ_{Source} synthesised by HRTFs recorded with a KEMAR head and torso simulator with 1° resolution in the horizontal plane. The HRTF database is described in [12] and is freely available for download. Fig. 1 shows the chosen source positions as the black loudspeakers, that coincided with the position of real loudspeakers in [1]. Therefore, the synthesised source directions stem from measurement of the loudspeaker positions and thus differ slightly from fig. 1. For source directions in between measured HRTFs, linear interpolation was applied. The distance induced level differences between the 11 sources were compensated for. The source content was 100 s of independent white noise pulses (pulse length 700 s with cosine-shaped 20 ms fade-in/fade-out, pause length 300 ms, bandpass filtered with 4th order Butterworth filter from 125 Hz–20 kHz) played back in a loop. The stimuli were the same as in the two previous studies [1, 11].

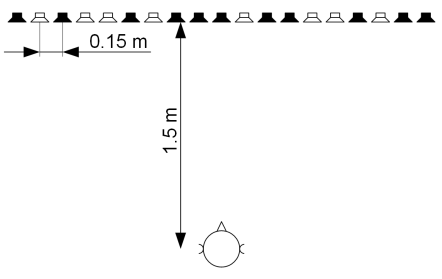

Figure 1: Location of virtual sources (loudspeaker symbols) and listener, only source positions in black are used in the listening experiment

Apparatus

The rendering of the stimuli in dynamic binaural synthesis over headphones (type AKG K601) was carried out by the SoundScape Renderer [13]. Head movements in azimuthal directions were provided by a head tracker type Polhemus Patriot. The sensor of this electromagnetic tracker was attached to the top of the headphones with the source about 1 m behind the head of the subject. The HMD type Oculus Rift CV1 provided a visual feedback of the head direction by an orange circle in a very simple spherical grid, cf. fig. 2. In the frontal direction only the horizon was visible except for the calibration phase where the 0° direction was marked to synchronise the independently running head tracking systems of the HMD and the electromagnetic tracker. Both devices were connected to a Windows system, so the timestamps for comparison were generated on the same machine. On a Linux system, the software for executing the listening test received the data from the electromagnetic tracker over network and passed it on to the rendering software. The sensor of the HMD was placed right in front of the subject. The simplified visual environment was chosen to avoid possible anchor effects, where subjects associate the direction of an auditory event with a prominent visual mark, like e.g. additional grid lines.

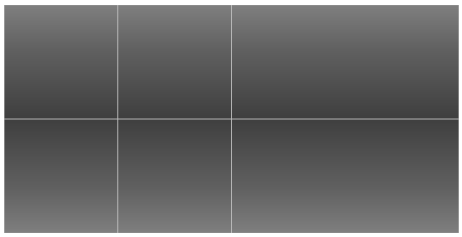

Figure 2: Image for simple spherical grid as visual VR, frontal direction in centre of right half.

Procedure

The subjects sat on a revolving chair wearing the headphones and the HMD with a keypad in their hands. After the calibration phase for the head trackers which included adjusting the HMD to the individual's head, subjects were instructed to point their head in the direction of an auditory event ignoring the vertical dimension. It was possible for subjects to complete this movement by both turning the head as well as turning on the revolving chair. Subjects were encouraged to perform oscillating head movements to help determining the direction. When the subjects found the direction of the auditory event, they pressed a button on the keypad and the mean of the last 5 values of the Polhemus Patriot tracking data were saved as result. After pressing the button, the next trial started. The 11 conditions had to be repeated 5 times by each subject leading to 55 trials presented in a randomised order with a preceding training of 11 trials (each condition once) also in randomised order.

Test subjects

10 subjects with an average age of 33 years participated in the listening test. 6 had home or professional experience in the field of audio, 8 had participated in listening tests before.

Results

Comparison of the two tracking systems

Fig. 3 compares the data of the two independent tracking systems with an exemplary movement of a human head ranging from approx. +90° to −90° azimuth while resting in between. As can be seen, there exist slight differences between the tracking systems that become larger for lateral head directions. The Oculus Rift tracking data appears to be a bit smoother presumably caused by the incorporated forecast of the trajectory, which can also lead to slight overshoots for more abrupt movements. The differences between the two systems in the azimuth range of the presented virtual sources between ±42° do not exceed 2°, which was tolerated in this listening test.

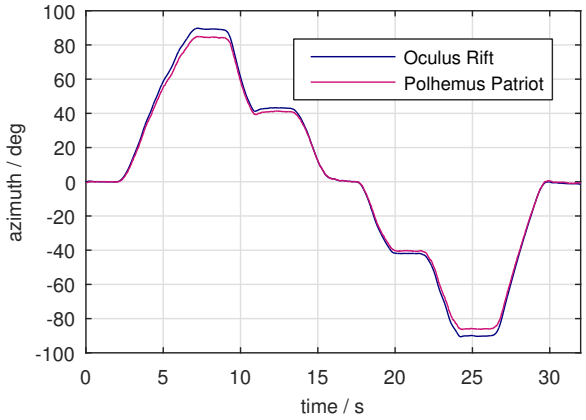

Figure 3: Comparison of azimuthal data of the two utilised tracking systems

Localisation error

The localisation errors as differences between the directions of the synthesised sources and the reported auditory events of the subjects are depicted in the histogram in fig. 4, left. The localisation error appears to follow a normal distribution, as was also the case for the results of the previous studies [1, 11], which allows for parametric statistic tests to be performed. Fig. 5 shows an overview of the results in comparison to the previous studies.With respect to the localisation errors, the present study shows with means not exceeding ±1.1° good results compared to the previous studies. To evaluate the localisation error depending on condition, a test on difference of the slope b of each regression line to zero is performed with the test statistic [14, Ch. 11.2.2]

$$t = b \cdot \sqrt{\sum_i (\phi_i - \bar{\phi})^2} \cdot \sqrt{\frac{n-2}{\sum_i (e_i - \hat{e}_i)^2}}, \qquad (1)$$

that follows Student's t-distribution with $n-2$ degrees of freedom. In eq. (1), (ϕ_i, e_i) are the mean localisation errors over condition (black markers in fig. 5), $\bar{\phi}$ is the mean source direction, $\hat{e}_i$ are the values predicted by the regression line and n is the number of data points. With Bonferroni correction for multiple testing, the null hypothesis $H_0 : \beta = 0$ (β: true slope of the population) can only be rejected for the results of [1], with a confidence level of 95%. This shows that the apparatus from [1] is

provoking undershoots of the reported answers for more lateral source directions while the present study and [11] do not suffer from this limitation. It has to be noted, that the coefficient of determination for the regression lines of [11] and the present study are quite low, as these lines are almost horizontal, thus representing an independence of the localisation error from the source direction. This must lead to a low correlation coefficient even if the horizontal regression line is optimal in a least-squares sense.

To compare the standard deviations of the three studies, a Bartlett test on homogeneity of variances is performed on the overall standard deviations $s_{[1]} = 4.8°$, $s_{[11]} = 3.1°$ and $s_{\text{present study}} = 4.2°$. The test confirms differences between the studies ($\alpha = 0.05$) and a follow-up pairwise F-test with Bonferroni correction for multiple testing reveals the following ranking of the standard deviations with a confidence level of 95%: $s_{[1]} > s_{\text{present study}} > s_{[11]}$. It has to be noted that the results of the previous studies have been corrected for the bias of each listener because of a possible bias introduced by the positioning of the laser pointer as it was not possible to mount it on the headphones so that it pointed exactly in the frontal direction. This data correction has not been performed on the results of the present study.

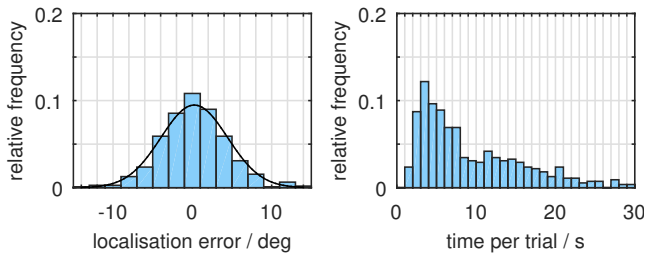

Figure 4: Normalised histograms for: left: localisation error compared to normal distribution with same mean and variance, right: elapsed time per trial

Elapsed time per trial

Subjects needed on average 9 minutes to complete the whole experiment (without training), minimum and maximum durations ranged from about 3.5 up to 20 minutes. Fig. 4, right, shows a histogram of the elapsed times per trial in s and table 1 shows a descriptive statistic compared to the results from previous studies.

Table 1: Statistics on elapsed time per trial in s

elapsed time	study in [1]	study in [11]	present study
arithmetic mean	5.6	9.0	9.8
5th percentile	2.2	2.6	2.3
median	4.6	6.8	7.1
95th percentile	13.6	24.2	24.7

Discussion

Humans are able to localise real broadband sound sources with an accuracy of up to 1° in the frontal horizontal plane [15]. For virtual sources based on non-individual HRTFs, this accuracy can be degraded [3]. A reporting method for the localisation of (virtual) auditory events should be at least as accurate as human localisation, which is the case for the presented apparatus with a maximum localisation error of 1.1°. Also the standard deviation appears to be acceptable compared to the previous studies and to the averaged median-to-quartile distance of 2.9° in [3] achieved for non-individual, but pre-selected HRTFs with the Proprioception Decoupled Pointer method (9 test subjects).

The independently working tracking systems exhibited deviations for more lateral directions as shown in fig. 3. Though this could have been a source of error, there is no dependence of the mean localisation error on the source direction. This appears also not to be the case for the standard deviation, cf. fig. 5.

The performance of the individual test subjects differed considerably in time devoted to solving the task and in standard deviations of the localisation error. There seemed to be no obvious relation between these two observations, though. Also, the performance of the subjects did not seem to depend on previous experience in listening tests or the field of audio. Possibly, a longer training phase as investigated by [9] is necessary to familiarise all subjects with the unusual task in a virtual environment with an HMD. The unfamiliar task could also be the reason for the medium standard deviation compared to the two previous studies. It has to be noted, though, that the data correction introduced by [11] is not only compensating the bias caused by inaccurate positioning of the laser pointer, but also any other source of bias, e.g. a test subject with a slight hearing loss on one ear, a randomly occurring bias or bias due to non-individual HRTFs, thus decreasing the measured standard deviation.

Conclusions

The presented reporting method for localisation tasks has been shown to be accurate enough for localisation experiments with virtual sources exhibiting a maximum localisation error in the frontal horizontal plane of about 1°. Varying performances of individual subjects suggest a need for an intensified training phase. The method can also be applied for experiments with elevated sources, but an investigation of the optimal visual environment with a trade-off between orientation marks and potential visual anchor effects should be performed first.

References

[1] Wierstorf, H., Spors, S., Raake, A.: Perception and evaluation of sound fields. Proc. of the Open Seminar on Acoustics, 2012

[2] Lewald, J., Dörrscheidt, G. J., Ehrenstein W. H.: Sound localization with eccentric head position. Behavioural Brain Research 108 (2000), 105–125

[3] Seeber, B.: Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode. Dissertation, Technical University of Munich, 2003

[4] Bahu, H., Carpentier, T., Noisternig, M., Warusfel, O.: Comparison of Different Egocentric Pointing

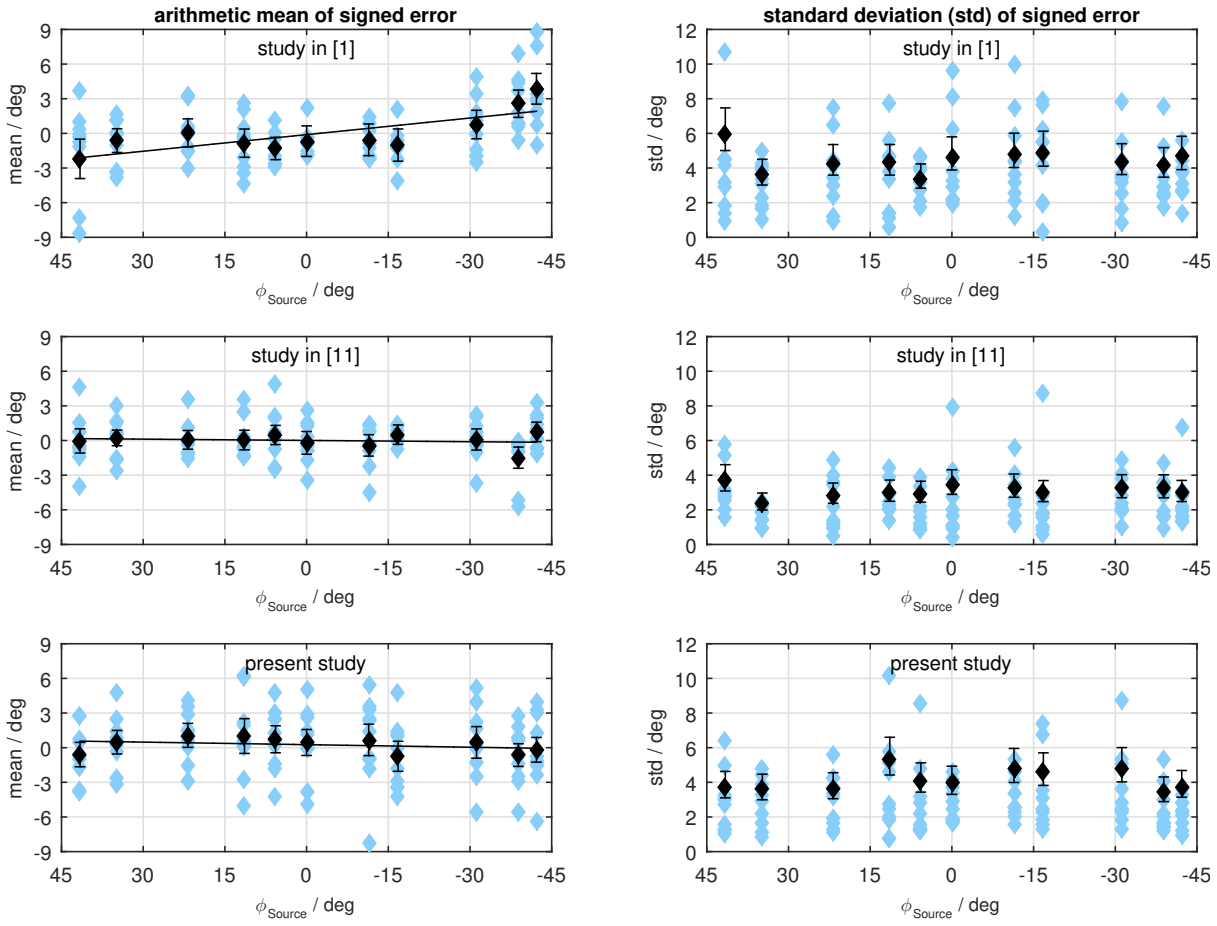

Figure 5: The rows show the arithmetic mean (left column) and standard deviation (right column) of the signed error for the three compared studies. Blue markers show results of individual subjects (10 subjects per study), black markers show results for all subjects together with the 95% confidence interval. The black lines in the left column represent the regression lines for the localisation errors.

Methods for 3D Sound Localization Experiments. Acta Acust. united Ac. 102 (2016), 107–118

[5] Schoeffler, M., Westphal, S., Adami, A., Bayerlein, H., Herre, J.: Comparison of a 2D- and 3D-based graphical user interface for localization listening tests. Proc. of the EAA Joint Symposium on Auralization and Ambisonics, 2014

[6] Frank, M., Mohr, L., Sontacchi, A., Zotter, F.: Flexible and intuitve pointing method for 3D auditory localization experiments. Proc. of the 38th AES Conference on Sound Quality Evaluation, 2010

[7] Bronkhorst, A. W.: Localization of real and virtual sound sources. J. Acoust. Soc. Am. 98 (1995), 2542–2553

[8] Spors, S., Schleicher, R., Jahn, D., Walter, R.: On the Use of Eye Movements in Acoustic Source Localization Experiments. Proc. of the 36th German Annual Conference on Acoustics (DAGA), 2010

[9] Majdak, P., Goupell, M. J., Laback, B.: 3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training. Attention, Perception, & Psychophysics 72 (2010), 454–469

[10] Pelzer, S., Kohnen, M., Vorländer, M.: Evaluation of Loudspeaker-based 3D Room Auralizations using Hybrid Reproduction Techniques. Proc. of the 40th German Annual Conference on Acoustics (DAGA), 2014

[11] Winter, F., Wierstorf, H., Spors, S.: Improvement of the reporting method for closed-loop human localization experiments. Proc. of the 142nd AES Convention, 2017

[12] Wierstorf, H., Geier, M., Raake, A., Spors, S.: A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances. Proc. of the 130th AES Convention, 2011

[13] Geier, M., Spors, S.: Spatial Audio with the SoundScape Renderer. Proc. of the 27th Tonmeistertagung – VDT International Convention, 2012

[14] Bortz, J., Schuster, C.: Statistik für Human- und Sozialwissenschaftler. Springer, Berlin Heidelberg, 2010

[15] Blauert, J.: Spatial Hearing. The MIT Press, Cambridge London, 1997