Influence of head tracking on distance estimation of nearby sound sources

Christoph Pörschmann[1], Johannes M. Arend[1,2], Philipp Stade[1,2]
[1] *TH Köln, Institute of Communications Engineering, Cologne, Germany*
[2] *TU Berlin, Audio Communication Group, Berlin, Germany*
Email: christoph.poerschmann@th-koeln.de

Introduction

Binaural synthesis serves as a powerful tool for psychoacoustic research, virtual environments or architectural acoustics. For many of those applications head tracking can be applied and the auralization be adapted accordingly to the listener's head orientation. By this, several aspects perceiving the auditory scene are enhanced. Many studies (e.g. [1][2]) have shown that head tracking can increase localization accuracy and significantly decrease front-back confusions. As shown in [3] appropriately considering head movements can as well result in a better externalization of sound sources. However, no relationship between head tracking and externalization was found e.g. in other studies, e.g. [4].
In this paper we analyze, if distance perception of nearby sound sources is affected by head tracking. The underlying question of our research is to investigate if monaural or binaural directional cues have an influence on distance perception in the near field. Brungart et al. [5] investigated the specific properties of sound incidence from nearby sound sources in great detail. We confirmed most of these properties in own measurements [6].
One prominent feature of near-field HRTFs (Head-Related Transfer Functions) is the increase of ILDs (Interaural Level Differences) as a function of source proximity. According to Brungart [5], especially at source distances below 0.5 m, this rise of ILDs is dramatic. Caused by frequency-dependent head-shadowing effects, the ILDs increase the more the sound source is positioned laterally to the head. This effect is stronger for closer distances; at 0.25 m the HRTFs showed ILDs of up to about 23 dB [6]. Thus ILDs might provide a relevant cue for distance perception in the proximal region. The ITDs (Interaural Time Differences) are as well maximal for sound incidence from about $\varphi = 90^\circ$ and $\varphi = 270^\circ$ in the horizontal plane. However, the ITDs are barely influenced by source distance.
Another prominent effect is the low-pass filtering character of nearby sources, meaning that sound sources are getting darker in timbre as they approach the head. This effect is perceived strongest for very close distances and sound sources at the front or rear. It might as well serve as a monaural cue for distance estimation here. All these cues are altered significantly when the head is turned and thus give the listener additional information. If directional cues (monaural and binaural ones) influence distance estimation of nearby sound sources, it can be expected that head movements are relevant here.
Scientific publications investigating the influence of head movements on distance estimation are rare. Simpson and Stanton [7] as well as Rosenblum et al. [8] performed experiments with a loudspeaker-based setup and investigated their influence on distance perception of nearby sound sources. Both studies revealed no influence of head movements on perceived source distance.
In order to investigate distance estimation in headphone-based virtual environments, specific near-field HRTFs are required which consider the acoustical specifics of nearby sound sources. In [6] we presented measurements of such near-field HRTFs and performed a first listening experiment which investigated if those near-field HRTFs can be applied to code distance. In this experiment we considered head tracking in all tested conditions and investigated if appropriate distance estimation is still possible when natural level differences between the stimuli are missing. For level-normalized stimuli we found no correlation between the chosen near-field HRTF set and the estimated distance. These findings are supported by the results from a comparable experiment without head tracking by Shinn-Cunningham et al. [9][10] who could as well not reveal any relationship between binaural cues and distance estimation in the near field. These results are in contrast to Brungart [11] who found that distance perception is still possible when loudness differences are removed. Brungart concluded that binaural cues are relevant for distance estimation of nearby sound sources in the proximal region.
Several other studies on distance perception of nearby sound sources were applied in virtual acoustic environments without head tracking: Parseihian et al. [12] compared distance perception of real and virtual sound sources. Kan et al. [13] as well as Spagnol et al. [14] used Distance Variation Functions (DVFs) in order to model the influence of nearby sound sources on the monaural and binaural cues. While Kan et al. [13] found improvements in distance estimation by using DVFs, the studies of Spagnol et al. [14] revealed that intensity cues overshadow any other cue and thus the DVFs do not allow a better distance estimation than intensity cues only. All these studies investigate if directional cues (monaural and binaural ones) have an influence on distance perception. This is the general question which motivated us to perform this study.
In the experiment presented here we used near-field HRTF datasets of a Neumann KU100 dummy head [6]. These datasets, which are freely available for download, contain impulse responses, measured spherically with a high angular resolution at sound source distances between 0.50 m and 1.50 m. Additionally HRTFs on a circular grid at distances between 0.25 m and 1.50 m were cap-

tured. In this paper we present preliminary results of one part of a series of experiments. In the same way as in [6] the basic task was to estimate auditory distance to a virtual sound source. This study investigated if head tracking affects distance estimation of nearby sound sources. The test was split in two groups: One group performed the tests with and the other without head tracking. It is important noting that we did not train the subjects before the experiment. Thus, they had to rely on their life experience in perceiving near-field sound sources.

Methods

Participants

Four females and 26 males aged between 21 and 28 years (M = 23.7 years, SD = 2.19 years) participated at this stage of the experiment (deadline 2016/10/31). Most of them were students in media technology or electrical engineering at TH Köln. Thirteen participants already took part in previous listening experiments and thus were familiar with the binaural system. None of the subjects reported any hearing problems.

Setup

The experiment took place in the anechoic chamber at TH Köln, which ensured a low background noise level of less than 20 dB(A). The experiment was implemented, controlled, and executed with the MATLAB-based software Scale [15], which also accessed the SoundScape Renderer [16] for binaural rendering. To acquire horizontal head movements, a Polhemus Fastrak was used. Vertical or translational head movements were disregarded. The subjects entered their responses on a tablet computer (iPad). The audio signal was presented over AKG K-601 headphones. Headphone compensation was applied according to [17] in order to equalize the binaural chain.

Materials

The anechoic test signal was a pink noise burst sequence with a burst length of 1500 ms (including 10 ms cosine-squared onset/offset ramps) and pauses of 500 ms. For the listening experiment, we used the circular grid measurements for all five distances from 0.25 m to 1.50 m. Per distance, we tested for three different sound incidence angles ($\varphi = 30°$, 150° and 270°). The playback level was set to 61 dB(A) Leq for stimuli simulating a sound source at a distance of 1 m. This resulted in a maximum playback level of about 79 dB(A) Leq for the closest distance of 0.25 m ($\varphi = 270°$).

Procedure

The participants had to rate distance on a seven-point category scale ("very close", "close", "rather close", "medium", "rather distant", "distant", "very distant"); a scale that had been successfully used in earlier experiments, e.g. [18]. The subjects were allowed to rate interim values between the given categories. The test procedure was as follows. The subjects were divided into two groups. One group performed the test without adapting the sound field to the listeners' head movement, for the other group head movements were appropriately considered.

As already mentioned above, there were no training session and no scale anchoring process. Informal pretests showed that training involves strong learning effects. For level-normalized stimuli we observed the following: In the beginning test persons could hardly distinguish between distances, but after they were given feedback, they learned to differentiate based on spectral changes, varying ILDs and head movement. However, we wanted to know if distance perception in the near field works instantaneously without prior knowledge about the auditory scene. Therefore, in the complete series of experiments we only gave basic instructions about the general procedure and the rating scale.

In each session, every participant was presented the five measured distances (0.25 m, 0.50 m, 0.75 m, 1.00 m, 1.50 m) for three different source azimuths ($\varphi = 30°$, 150°, 270°). For each trial, a user interface was displayed on the tablet computer showing five value faders ranging from "very close" to "very distant" (see Figure 1). The five faders corresponded to the five actual measured distances, thus the subjects had to rate multiple stimuli per trial. The source azimuth was the same for all distances (or faders) within a trial. By touching the respective fader, the participants were able to switch between the corresponding stimuli as often as required. Technically speaking, the HRTF filter-set was switched when touching the fader while the noise sequence was played in a loop. The order of the faders per trial as well as the order of the trials itself were randomized.

The procedure was repeated 10 times per azimuth, thus a full run consisted of 30 trials (with five distance ratings per trial). Regardless of whether the subjects were presented the head-tracked or the not head-tracked stimuli they were encouraged to move their head during the estimation process in the form of (small) localization movements. However, they had to keep their front viewing direction because of the different source directions. In total, the test lasted for about one hour including the verbal instruction, a short break, and three post-experiment questions.

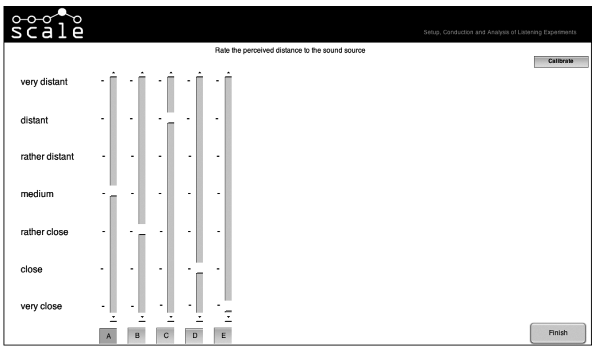

Figure 1: User interface of the experiment. On the left, the seven-point category scale is displayed. The five faders correspond to the five presented distances, randomly ordered for each trial.

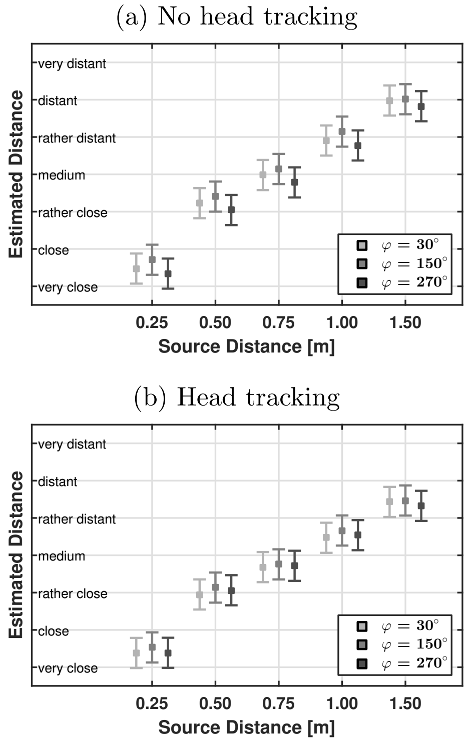

Figure 2: Mean estimated distances for the group without head tracking (a) and with head tracking (b) as a function of source distance (abscissa) and source azimuth (colors). The error bars denote 95% within-subject confidence intervals based on the respective main effect of distance.

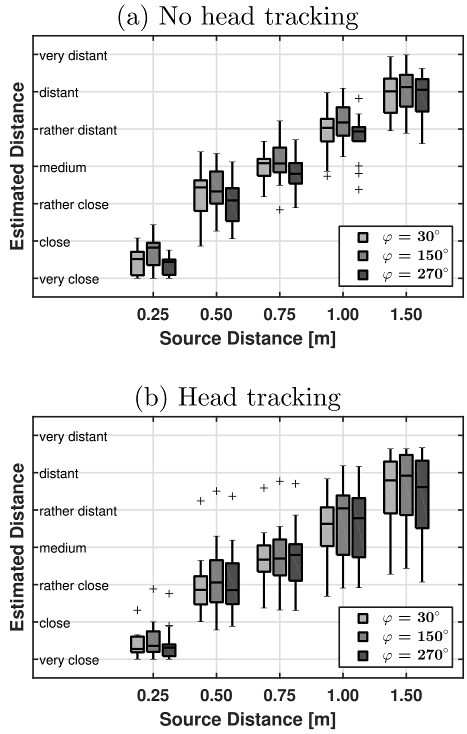

Figure 3: Distribution of distance ratings for the group without head tracking (a) and with head tracking (b) as a function of source distance (abscissa) and source azimuth (colors). The plots show the median and the quartiles as well as the outliers (larger than 1.5 times the interquartile range).

Results

The following statistical analysis is based on the mean value per subject, thus the 10 trials per subject for each condition were averaged first. We performed a mixed ANOVA with the between-group factor "head tracking" and the within-subjects factors "distance" and "azimuth". A Greenhouse-Geisser (GG) correction [19] (for tests with more than one degree of freedom in the numerator, where GG is appropriate) was conducted for the within-subject effects. The ANOVA showed no significant influence of head tracking. ($F(1, 28) = 1.63$, $p = .212$, $\eta_p^2 = .05$). Thus no significant differences between the group of subjects who performed the test either with head tracking or without were observed. As expected we found a significant main effect of "distance" ($F(4, 112) = 435$, $p < .001$, $\eta_p^2 = .94$, $\epsilon = .37$). Furthermore, a main effect of "azimuth" ($F(2, 56) = 15.6$, $p < .001$, $\eta_p^2 = .36$, $\epsilon = .88$) was observed. Only one small significant interaction effect for "head tracking" x "azimuth" ($F(2, 56) = 4.54$, $p = .019$, $\eta_p^2 = .14$) was found. This interaction effect describes direction-dependent differences between the head-tracking and the non head-tracking condition. As shown in Figure 2 the differences in distance estimation which can be observed for the different incidence directions without head tracking diminished when head tracking was applied. Figure 2 presents the respective means of estimated distance for the conditions with and without head tracking, averaged over subjects. The error bars display 95% within-subject confidence intervals [20], based on the error term of the distance main effect. In Figure 3 the distribution of the same data in form of boxplots is shown. Surprisingly the results vary more for the conditions with head tracking. In a next step we investigated the subjects' head movements. We measured the viewing direction during the complete experiment (one value per second). However, for one subject in the non head-tracked condition the data was not collected appropriately. Thus for this condition results for only 14 out 15 subjects are considered. In Figure 4 the distributions of the viewing directions are shown. Without head tracking the SD was 13.4° while head tracking increased the SD to 18.4°. We tested the inter-subject differences of the SD in a t-test for both groups which revealed a significant effect of head tracking ($t(27) = 2.26$, $p = .03$).

Conclusion

The presented investigation was carried out as a part of a broader study on distance perception of nearby sound sources. We investigated if head movements have an influence on distance perception of nearby sound sources for untrained listeners. The results revealed no significant influence of head tracking on estimated distance, even though participants moved the head more when head tracking was activated. Surprisingly the results varied more for the head-tracked conditions than for the not head-tracked ones. Our results support the findings of [7] and [8] who with a loudspeaker-based setup as well could not reveal an influence of head movements on distance es-

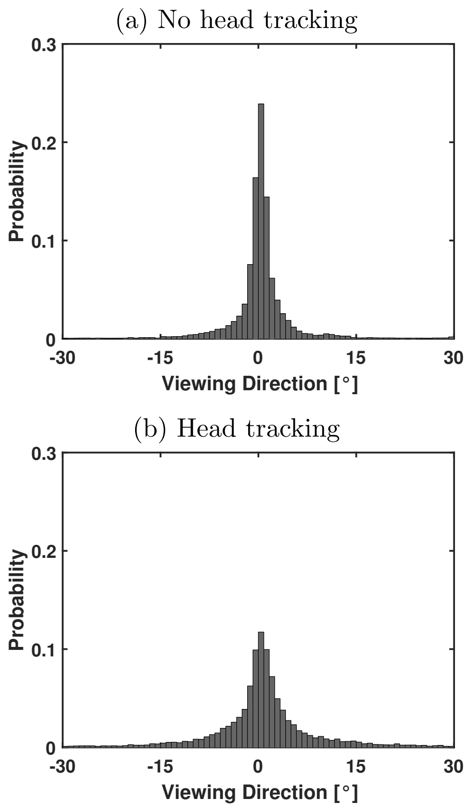

Figure 4: Distribution of the viewing directions for not head-tracked condition (a) and head-tracked condition (b). The viewing directions were averaged over all subjects, distances and incidence directions. It can be observed that the head movements are increased in the head-tracked scenarios.

timation. Generally this study supports the thesis that monaural and binaural directional cues are of minor relevance for distance perception of nearby sound sources. Several scientific relevant questions still remain open. Future investigations could address the influence of the training of the listeners on the results and it could be addressed which other perceptual attributes of nearby sound sources are influenced by head tracking. Finally it is worth noting that in our studies no translational movements were regarded which especially for nearby sound sources have a strong influence on the ear signals.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) under the support code 03FH014IX5-NarDasS. We thank all participants of the listening experiment. Great thank to Tim Lübeck and Patrick Pereira for their support during the experiments.

References

[1] Blauert, J., *Spatial Hearing - Revised Edition: The Psychoacoustics of Human Sound Source Localisation*, MIT Press, Cambridge, MA, 1997.

[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," *Acta Acustica united with Acustica*, 86, pp. 1046–1053, 2000.

[3] Brimijoin, W. O., Boyd, A. W., and Akeroyd, M. A., "The contribution of head movement to the externalization and internalization of sounds," *PLoS ONE*, 8(12), pp. 1–12, 2013.

[4] Begault, D. R., Wenzel, E. M., and Anderson, M. R., "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source." *Journal of the Audio Engineering Society*, 49(10), pp. 904–916, 2001.

[5] Brungart, D. S. and Rabinowitz, W. M., "Auditory localization of nearby sources. Head-related transfer functions," *Journal of the Acoustical Society of America*, 106(May), pp. 1465–1479, 1999.

[6] Arend, J. M., Neidhardt, A., and Pörschmann, C., "Measurement and Perceptual Evaluation of a Spherical Near-Field HRTF Set," *Proceedings of the 29th Tonmeistertagung - VDT International Convention*, 2016.

[7] Simpson, W. E. and Stanton, L. D., "Head Movement Does Not Facilitate Perception of the Distance of a Source of Sound," *American Journal of Psychology*, 86(1), pp. 151–159, 1973.

[8] Rosenblum, L., Wuestefeld, A. P., and Anderson, K. L., "Auditory reachability: An affordance approach to the perception of sound source distance," *Ecological Psychology*, 8(1992), pp. 1–24, 1996.

[9] Shinn-Cunningham, B., Santarelli, S., and Kopco, N., "Distance Perception of Nearby Sources in Reverberant and Anechoic Listening Conditions : Binaural Vs . Monaural Cues," 1999.

[10] Shinn-Cunningham, B. G., "Localizing sound in rooms," *ACM/SIGGRAPH and Eurographics Campfire: Acoustic Rendering for Virtual Environments*, (May), pp. 17–22, 2001.

[11] Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M., "Auditory localization of nearby sources. II. Localization of a broadband source." *Journal of the Acoustical Society of America*, 106(October), pp. 1956–1968, 1999.

[12] Parseihian, G., Jouffrais, C., and Katz, B. F. G., "Reaching nearby sources: Comparison between real and virtual sound and visual targets," *Frontiers in Neuroscience*, 8(SEP), pp. 1–13, 2014.

[13] Kan, A., Jin, C., and van Schaik, A., "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function." *Journal of the Acoustical Society of America*, 125(4), pp. 2233–42, 2009.

[14] Spagnol, S., Tavazzi, E., and Avanzini, F., "Distance rendering and perception of nearby virtual sound sources with a near-field filter model," *Applied Acoustics*, 115, pp. 61–73, 2017.

[15] Vazquez Giner, A., "Scale - Conducting Psychoacoustic Experiments with Dynamic Binaural Synthesis," *Proceedings of the DAGA2015*, pp. 1128–1130, 2015.

[16] Geier, M., Ahrens, J., and Spors, S., "The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods," in *Proceedings of 124th Audio Engineering Society Convention 2008*, pp. 179–184, 2008.

[17] Bernschütz, B., "A Spherical Far Field HRIR / HRTF Compilation of the Neumann KU 100," *Proceedings of the DAGA 2013*, pp. 592–595, 2013.

[18] Pörschmann, C. and Störig, C., "Investigations into the velocity and distance perception of moving sound sources," *Acta Acustica united with Acustica*, 95(4), pp. 696–706, 2009.

[19] Greenhouse, S. W. and Geisser, S., "On methods in the analysis of profile data," *Psychometrika*, 24(2), pp. 95–112, 1959.

[20] Loftus, G. R. and Masson, M. E. J., "Using confidence intervals in within-subject designs," *Psychonomic Bulletin and Review*, 1(4), pp. 476–490, 1994.