Investigating the Effect of Head Movement on the Perception of Reproduction Artefacts of Moving Sources

Matthieu Kuntz, Bernhard U. Seeber

Audio Information Processing, Technical University of Munich

E-Mail: {matthieu.kuntz; seeber}@tum.de

Introduction

The use of virtual acoustics in psychoacoustic and hearing research is increasingly widespread. The ability to place virtual sources at arbitrary locations increases the angular resolution beyond that of the physical loudspeakers and allows the synthesis of moving stimuli. Being able to recreate realistic acoustic scenes in laboratory conditions without the need for headphones also enables the use of virtual acoustics for hearing aid or cochlear implant users [1]. A previous study by the authors [2] investigated the auralization of moving sound sources using combinations of Higher-Order Ambisonics at low frequencies and other panning techniques at high frequencies, in order to minimize the sound pressure level errors in and around the center of the loudspeaker array [3].

When designing increasingly realistic acoustic environments and listening experiments, it is crucial to allow participants to turn their head freely. For moving sources in the foreground, it is likely that at least some part of the trajectory is followed with a head turn. The key differences when following a sound source with head turns compared to keeping a constant head orientation are the different interaural cues and the head above torso orientation, introducing audible changes in the HRTFs [4], and affecting localization performance [5].

Given a higher sensibility to interaural cue changes in the front [6,7], a higher sensibility to reproduction artefacts could be expected. A head rotation also introduces a translation of the ear, which causes a change in the high frequency comb filters observed at off-center positions in loudspeaker arrays with time of flight equalization [8]. On the other hand, focusing on a sources' movement might introduce higher cognitive overhead, leading to a less pronounced artefact perception. This study compares ratings of perceived artefacts in the reproduction of a moving source for different panning techniques and head movement conditions.

Methods

Experimental setup

The experiment took place in the Simulated Open Field Environment (SOFE) loudspeaker array in the anechoic chamber of the Professorship for Audio Information Processing of the Technical University of Munich [9]. Participants were seated on a chair in the center of the loudspeaker array, facing the 0° direction. Four participants had their head tracked in the experiment.

Stimuli

The stimulus was a pink noise source between 100 Hz and 15 kHz, with a Gaussian envelope to create a 50 ms fade-in and fade-out. The source was moving in a circular trajectory at constant angular speed, which was varied between 10°/s, 20°/s, 30°/s, 60°/s and 90°/s across the experiment. The source was moving from -30° to 30° or vice-versa, the position being defined at the 67.5% point of the fade-in and fade-out envelopes. Since combining HOA with other panning techniques at high frequencies increased the perceived artefacts compared to a HOA only auralization [2], this study did not consider those combinations. The stimuli were played back in 2D over the 36-channel loudspeaker array with 10° angular spacing between loudspeakers.

Panning techniques

This study investigates four different panning techniques, which are briefly listed here. For more details, refer to our previous publication [2]. In nearest loudspeaker mapping (NLS), the source position is played back from the nearest loudspeaker, which introduces an average position error of 2.5° for the 10°-spaced loudspeakers of the array. Vector base amplitude panning (VBAP, [10]) uses two loudspeakers closest to the sound source position and determines their gains based on the source position between them. Perceptually equalized panning (PEP, [11]) uses a spherical head model to compute short FIR filters to correct for the angle dependent HRTF difference between a virtual source and its loudspeaker-based reproduction, which are applied to VBAP. The loudspeaker driving functions were not changed or adapted to the head movements of the participants and were computed with the assumption of a static listener facing 0°. Higher-Order Ambisonics (HOA) was implemented with the *basic* or *sampling* decoder, as described in [2,8].

Procedure

The experiment was split into two parts. In the first part, participants were asked to keep their head fixed and look towards the front. A stimulus with random direction and speed was played back, for which the participants were asked to rate the dynamic artefacts they heard on a scale of 1 (not audible) to 7 (extremely audible). The confirmation of a rating triggered the next stimulus playback. Participants underwent a familiarization run to get an overview of the different presentation conditions. The first part was split into 10 runs of around 10 minutes each.

For the second part of the experiment, participants were instructed to rotate their head and point their noise towards the sound source location. The presented stimuli were grouped by speed and the movement direction was alternated from trial to trial, resulting in a predictable source trajectory. Participants used the same rating scale as in the first part. The second part was split into 12 runs of around 10 minutes, each run starting with a short re-familiarization session with a static head to allow participants to remember their rating scheme from the first part of the experiment. After that, participants were instructed to turn their heads. 16 dummy trials were run to give participants time to learn the head turns to follow the source, reducing the cognitive effort to follow the source to a minimum.

Eight normal hearing subjects participated in both parts of the experiment. Their threshold in quiet was measured before the experiment and lied below 20 dB HL for frequencies between 250 Hz and 8 kHz. Each individual condition was repeated 7 times, resulting in total of 5 speeds × 4 panning techniques × 2 directions × 2 head movement conditions × 7 repetitions = 560 trials analyzed in this study.

Results

Effect of panning technique

Figure 1 shows the median dynamic artefacts ratings for each participant, grouped by panning method and by head movement condition.

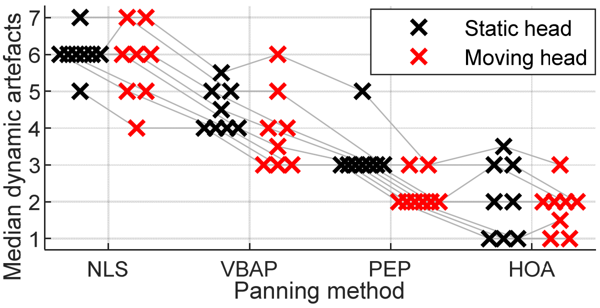

Figure 1: Median dynamic artefacts rating for each participant, grouped by panning method and head movement condition. Grey lines link the ratings of each participant.

When using NLS to reproduce a moving sound source, the results are not influenced by the head movement condition, the median rating across participants being 6 in both static and moving head conditions. For VBAP, the median ratings are slightly lower when participants followed the sound source. This decrease was observed in 6 out of 8 participants. The PEP method showed the highest decrease in ratings, visible for 7 out of 8 participants and showing similar results to HOA for the moving head condition. A slight decrease in median ratings was also observed in HOA.

Effect of source speed

Figure 2 shows the median dynamic artefacts ratings for each participant, grouped by source speed and by head movement condition. As the source speed increases, participants gave higher artefact ratings in the static head condition, median values increasing from 3 to 5. This was not observed in the moving head condition, where the ratings stayed almost constant across source speeds. This can also be seen in the increasing difference between static and moving head conditions with source speed.

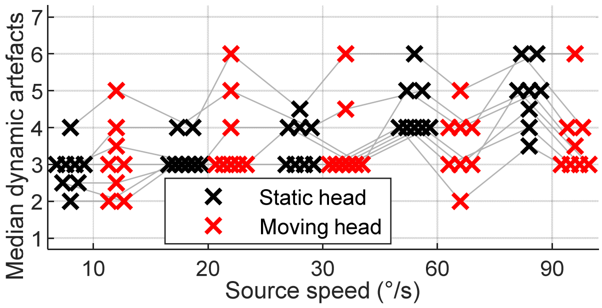

Figure 2: Median dynamic artefacts rating for each participant, grouped by source speed and head movement condition. Grey lines link the ratings of each participant.

Discussion

The overall ranking of the different panning techniques is preserved when listeners follow the sound source by turning their head: NLS was rated highest, followed by VBAP, PEP, and HOA. The ratings of NLS and HOA did not vary much between head movement conditions. In the case of NLS, this is probably due to a ceiling effect of the rating scale. The clearly discernable jumps in the supposedly smooth trajectory were often perceived as extreme, although the ratings decreased slightly with increasing source speed. The largest difference in ratings was observed for PEP, even though the FIR-filters were not recomputed to account for head turns, which should in theory yield higher reproduction errors. However, the filters in the front are very similar to one another, with differences of 0.5 dB or lower in the frequency range of 100 Hz – 15 kHz, so the errors introduced by turning the head towards the sound source are negligible.

We also observe an interaction effect between source speed and head movement condition. Higher velocities lead to higher artefact ratings in the static head case, a trend which is not observed in the moving head condition. At low source speeds, participants gave more extreme ratings, which did not affect their median rating. This could however be an indication of lower concentration ability for those trials. This was also reported by some participants, who mentioned that their ability to differentiate artefacts was lower and resulted in more bipolar distribution of responses. At higher speeds, the majority of trials were rated with 3 or 4, around the middle of the rating scale.

Conclusion

This work presents a listening experiment in which participants rated the amount of dynamic artefacts perceived in the auralization of a moving sound source. When participants were asked to turn their head to follow the sound source, the ratings were slightly lower compared to the trials where participants were asked to keep their head still. This effect was stronger for faster source movements. Further work should study the type of artefacts closer to determine what aspects of the sound drive the artefact ratings. Furthermore, the recorded head movements can be used to recreate the acoustic conditions participants observed in the moving head condition to investigate interaural cues changes the effect of self-motion on the artefact ratings.

Acknowledgements

The SOFE system was financed by the BMBF Bernstein Center for Computational Neuroscience, 01GQ1004B. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 352015383 – SFB 1330 C5.

Literature

[1] Kerber, S., Seeber, B.U.: Localization in Reverberation with Cochlear Implants: Predicting Performance from Basic Psychophysical Measures. Journal of the Association for Research in Otolaryngology 14(3) (2013), 379-392

[2] Kuntz, M., Seeber, B.U.: Investigating the Smoothness of Moving Sources Reproduced with Panning Methods. Fortschritte der Akustik (2022), 1465-1468

[3] Kuntz, M., Kolotzek, N., Seeber, B.U.: Gemessener Schalldruckpegel im Lautsprecherarray für verschiedene Schallfeldsyntheseverfahren. Fortschritte der Akustik (2021), 1437-1440

[4] Brinkmann F., Roden, R., Lindau, A., Weinzierl, S.: Audibility and Interpolation of Head-Above-Torso Orientation in Binaural Technology. IEEE Journal of Selected Topics in Signal Processing 9(5) (2015), 931-942

[5] Kolotzek, N., Gomez, G., Seeber, B.U.: The Effect of Head Turning on Sound Localization in the Horizontal Plane. Fortschritte der Akustik (2017), 1150-1153

[6] Yost, W.A.: Fundamentals of Hearing. Academic Press, San Diego, 2000

[7] Domnitz, R.: The interaural time jnd as a simultaneous function of interaural time and interaural amplitude, The Journal of the Acoustical Society of America 53 (1973), 1549-1552

[8] Zotter, F., Frank, M.: Ambisonics: A Practical 3D Audio Theory for Recording. Springer, 2019.

[9] Seeber, B.U., Clapp, S.: Interactive simulation and free-field auralization of acoustic space with the rtSOFE. The Journal of the Acoustical Society of America 141(5) (2017), 3974

[10] Pulkki, V.: Virtual Sound Source Positioning Using Vector Base Amplitude Panning. Journal of the Audio Engineering Society 45(6) (1997), 456-466

[11] Seeber, B.U., Hafter, E.: Perceptual equalization of artifacts of sound reproduction via multiple loudspeakers. Meetings on Acoustics (2013), 050045