

The importance of head movements for binaural room synthesis - a pilot experiment

Pauli Minnaar, Jan Plogsties, Søren Krarup Olesen, Flemming Christensen, Henrik Møller

Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B4, DK-9220 Aalborg, Denmark
acoustics@acoustics.auc.dk

DAGA 2000, Oldenburg, Germany, 20 - 24 March 2000

The impact of head movements on sound localisation of loudspeakers in a listening room was investigated. Listeners were allowed to move their heads during sound presentations in both real life and during binaural synthesis. The binaural synthesis was implemented by means of measured binaural room impulse responses, for which the direct sound was updated according to the head position of the listener. Localisation generally improved when head movements were allowed, compared to when listeners kept their heads still during sound presentations. This result encourages the use of head tracking in binaural synthesis systems.

1 Introduction

The aim of binaural technology is to provide at the eardrums of a person the sound pressures that would have been there in a listening situation. If this is done successfully the complete auditory experience is assumed to be reproduced. Since binaural technology considers air conduction to the eardrums as the only input to the auditory system, all information about an acoustical event (including spatial aspects) is contained in the two eardrum signals.

A distinction is made between 1) binaural recording, where the ear signals are recorded in a real environment and 2) binaural synthesis, where ear signals are created in correspondence with sound sources in a virtual environment. The latter is implemented by convolving an anechoic recording with a directional filter. If the directional filter is a head-related transfer function (HRTF) only the direct sound from a source to the ears of the listener is represented, whereas a binaural room impulse response (BRIR) includes the reflections from room boundaries as well.

HRTFs and BRIRs can be measured with small microphones in the ear canals of a person. Since this is inconvenient in practice a representation of a human head, called an artificial head, is more often used. In this work the artificial head VALDEMAR (developed at Aalborg University) was employed.

Localisation experiments are often used to determine the success of a binaural system. A listener (wearing headphones) is usually required to keep the head still during sound presentation. Both under anechoic conditions and in listening rooms the results are generally significantly poorer than in real life. This margin must be reduced substantially if binaural technology were to find widespread application.

Binaural technology has the unique ability of providing the full spatial information available to a person through only two audio channels. The listening experience is, however, limited if the listener's head is fixed in the recorded sound field. It may be argued that some principal localization cues are absent since the changes introduced to the ear signals, when moving the head, are not represented.

Since it is generally accepted that head movements improve sound localisation in real life a subjective experiment was done in a listening room to investigate the impact of head movements in binaural synthesis. Head movements were allowed in both real life and binaural synthesis and the results are compared to fixed-head localisation.

2 Previous work

A substantial amount of literature exists on sound localisation experiments with binaural recordings, where head movements were not allowed. Thorough reviews of 5 studies with human heads and of 18 studies with artificial heads are presented in Møller et al. [1] and [2] respectively.

Much fewer studies have been done where head movements were possible during binaural synthesis. However, Wenzel [3] demonstrated that enabling head tracking in binaural synthesis dramatically improves localisation even if nonindividualised HRTFs are used for synthesis. Sandvad [4] used individualised HRTFs for dynamic (head tracked) binaural synthesis. The results show localisation performance that is only slightly worse than in real life.

A binaural room synthesis system is described by Inanaga et al. [5], who showed that localisation can be done successfully using BRIRs. BRIRs were measured in a listening room for loudspeaker positions in the horizontal plane using the KEMAR artificial head. The study found that sound localisation with headphones is superior when head movements are allowed.

Horbach et al. [6] investigated the impact of head rotations on sound localisation in the horizontal plane. The Neumann KU100 artificial head was employed to measure both HRTFs in an anechoic chamber and BRIRs in a listening room. The localisation results with both HRTFs and BRIRs show less localisation inaccuracy as well as front/back confusions when head movements were possible.

Mackensen et al. [7] described a listening experiment where horizontal as well as vertical head movements were possible. The results show only a slight improvement when head tracking of both horizontal and vertical head movements was enabled, when compared to head tracking of horizontal head movements only.

In a localisation experiment by Begault et al. [8] the impact of head movements, reverberation and individualised HRTFs were compared. It was found that localisation accuracy is not significantly improved, although front/back reversals are almost completely eliminated by head tracking in a binaural synthesis system.

3 Methods

In the present study the same experimental set-up of loudspeakers in a standard listening room was used as described in Møller et al. [1, 2]. Nineteen ball loudspeakers were placed on stands in the room. Thirteen of the loudspeakers were located on the sphere around the listener's head at 1 m distance. The remaining loudspeakers were in front at distances of 1.7 m, 2.9 m and 5 m and 45° to the right of the listener at distances of 1 m, 1.7 m and 2.9 m. In a neighbouring control room a computer interfaced a relay unit that channelled a signal to any one of the loudspeakers.

The artificial head VALDEMAR was placed in the listener's position on a chair between the loudspeakers. A BRIRs was measured for every loudspeaker, using maximum length sequences. Furthermore, the HRTFs of the artificial head were measured in an anechoic room for 97 directions on the sphere around the head.

For every BRIR the direct sound was removed to obtain the later part of the impulse response that was kept constant for every loudspeaker. During binaural synthesis the head was tracked by means of a Polhemus FASTRAK and the direct sound was updated accordingly. This was implemented by filtering a dry signal separately with the HRTFs (updated with head movements) and the later part of the BRIRs (kept constant for each loudspeaker). The two signals were added and presented through a Beyerdynamic DT990 headphone. The headphone was equalised from measurements made on the artificial head.

The stimulus was a 5 s long female speech sentence, which was recorded in an anechoic room with the microphone at 300 mm distance from the mouth. This signal was input directly to one of the loudspeakers in the set-up during real life listening.

Listeners were required either to keep the head still or to move the head in the horizontal plane. When head movements were not allowed the listener was requested to look straight ahead. The head tracking was however active at all times during the binaural synthesis.

The listeners (2 male, 3 female) in the experiment were between 24 and 38 years old and none of them had participated in localization experiments before.

The experiment was automated and a small ‘traffic light’ prompted the actions of the listener. An answer was submitted by pressing with a pen on an electronic tablet holding a schematic drawing of the loudspeaker set-up. During a listening session 57 sounds (3 per loudspeaker) were presented. The sounds in a session were randomised individually for every listener and the order of the sessions was balanced to counteract the effects of learning.

4 Results

The results of this pilot listening experiment are shown in Figure 1. The errors made in localisation were decomposed into 4 categories. Sound sources that produce the same interaural time difference are positioned on an approximate cone called a cone of confusion. If a stimulus and a response are not on the same cone it is denoted an *out-of-cone* error. When errors are made by confusing directions on the same cone it is denoted a *within-cone* error, except if the stimulus and response are in the median plane, in which case it is a *median-plane* error. A response given in the same direction as the stimulus, but at an incorrect distance, is a *distance* error.

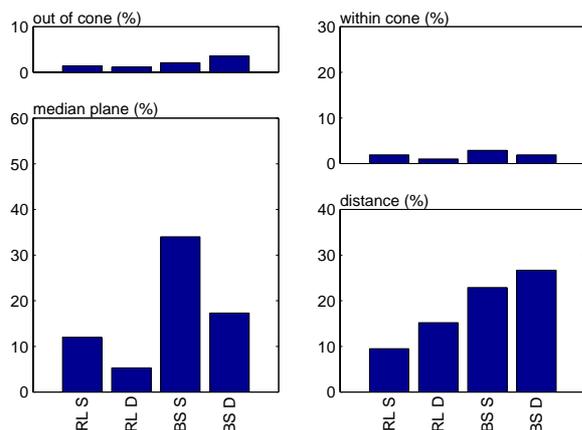


Figure 1. Localisation errors as a percentage of the potential number of errors in each category. The following abbreviations are used: RL real life, BS binaural synthesis, S static, D dynamic.

5 Discussion

The results show, for all error categories, that localisation in real life produces fewer errors compared to binaural synthesis. This is particularly clear in the median plane for static listening. When head movements are allowed the number of errors in both real life and binaural synthesis is reduced by approximately the same percentage. This implies, however, that the binaural synthesis has improved significantly (by almost 17 %).

The out-of-cone and within-cone errors are relatively few for both real-life and binaural synthesis, as well as for dynamic and static listening.

In contrast to the results in the median plane the number of errors made in distance perception are slightly higher for dynamic listening when compared to static listening. This is the case when listening both in real life and with binaural synthesis. It may be argued that the contribution of head movements to distance perception in a room is relatively small. Therefore, giving the listener the (slightly unnatural) task of moving the head in the horizontal plane only, may be more distracting than helpful for locating the active loudspeaker.

With the experiment described here it was shown that localisation in binaural synthesis generally improves when head movements are allowed, demonstrating the importance of dynamic auditory cues. Furthermore, in dynamic binaural synthesis the number of localisation errors is relatively low even though listeners’ own heads (and ears) were not used to implement the synthesis. The same result was found in previous studies [6, 7 and 8]. However, in those studies the complete BRIR was updated according to the head movements, whereas in this work only the direct sound had to be updated in real time, which lead to a substantially reduced computational load.

The listening experiment described here is considered as a pilot experiment that encourages further investigation into dynamic listening with binaural systems. The set-up will be used as a test platform where listeners can compare simulated (virtual) sound sources to the real sources in the same room.

6 Acknowledgements

We want to thank Claus Vestergaard Skipper for his competent assistance in the laboratory.

7 References

- [1] H. Møller, M. F. Sørensen, C. B. Jensen, D. Hammershøi, “Binaural technique: Do we need individual recordings?”, *J. Audio Eng. Soc.*, Vol. 44, No. 6, pp. 451-469, June 1996.
- [2] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sørensen, “Evaluation of artificial heads in listening tests”, *J. Audio Eng. Soc.*, Vol. 47, No. 3, pp. 83-100, March 1999.
- [3] E. M. Wenzel, “What perception implies about implementation of interactive virtual acoustic environments”, *Presented at the 101st Convention of the Audio Engineering Society*, Los Angeles, California, USA, November 1996 (preprint 4353).
- [4] D. R. Begault, E. M. Wenzel, A. S. Lee, M. R. Anderson, “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source”, *Presented at the 108th Convention of the Audio Engineering Society*, Paris, France, February 2000 (preprint 5134).
- [5] J. Sandvad, “Dynamic aspects of auditory virtual environments”, *Presented at the 100th Convention of the Audio Engineering Society*, Copenhagen, Denmark, May 1996 (preprint 4226).
- [6] K. Inanaga, Y. Yamada, H. Koizumi, “Headphone system with out-of-head localization applying dynamic HRTF (Head-Related Transfer Function)”, *Presented at the 98th Convention of the Audio Engineering Society*, Paris, France, February 1995 (preprint 4011).
- [7] U. Horbach, A. Karamustafaoglu, R. Pellegrini, P. Mackensen, G. Theile, “Design and applications of a data-based auralization system for surround sound”, *Presented at the 106th Convention of the Audio Engineering Society*, Munich, Germany, May 1999 (preprint 4976).
- [8] P. Mackensen, M. Fruhmann, M. Thanner, G. Theile, U. Horbach, A. Karamustafaoglu, “Head tracker based auralization systems: Additional consideration of vertical head movements”, *Presented at the 108th Convention of the Audio Engineering Society*, Paris, France, February 2000 (preprint 5135).