THE IMPORTANCE OF HEAD MOVEMENTS FOR BINAURAL ROOM SYNTHESIS

Pauli Minnaar, 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

ABSTRACT

The contribution of head movements to sound localisation of loudspeakers in a listening room was investigated. An experiment was done where listeners either were requested to keep their heads still while sound was being presented or were allowed to move freely while seated on a chair. The results give an indication of how well listeners can perform the task and what contribution head movements has.

The same localisation task was performed while listeners were presented with a simulation of the loudspeakers implemented by means of binaural synthesis. As in the “real life” case listeners either had to sit still or were allowed to move their heads. The binaural synthesis was implemented by means of measured binaural room impulse responses (BRIRs), for which the direct sound was updated according to the head position of the listener.

Pilot experiments showed that localisation generally improves when head movements are allowed, compared to when listeners keep their heads still during sound presentations. This encouraging result lead to the design of a full-scale experiment that is currently under way. The results will be reported at the ICAD2000 conference.

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 given listening situation. If the input to the auditory system corresponds to that which would exist due to a natural source with a certain location with respect to the listener the auditory experience should be the same independent of the true origin of the sound. Therefore headphones can be used to present sounds that appear to have any position with respect to the listener.

When a real event is recorded and reproduced it is referred to as binaural recording, whereas with binaural synthesis a world that might not exist in reality can be created. The fact that binaural synthesis only requires two audio channels to represent the entire three-dimensional space makes it particularly applicable for an auditory display.

Binaural synthesis still needs to be improved, however. Localisation experiments are often used to determine the success of a binaural system. Both under anechoic conditions and in listening rooms the results are generally significantly poorer than in real life. This margin must be reduced substantially before binaural technology will find widespread application.

It is recognised that in most previous experiments a listener (wearing headphones) is required to keep his/her head still during sound presentation. The listening experience is, therefore, limited since the listener’s head is fixed in the recorded sound field. It may be argued that some principal localisation cues are absent if the changes introduced to the ear signals, when moving the head, are not represented.

A substantial amount of literature exists on sound localisation experiments with binaural recordings or synthesis, where head movements were not allowed. Much fewer studies have been done where head movements were possible during binaural synthesis. These studies generally show, however, that localisation improves when head movements are allowed.

Furthermore, in most studies only the direct sound is implemented in the binaural synthesis. This can be done by convolving an anechoic recording with a head-related transfer function (HRTF). However, all the acoustical information related to the listening environment can be represented by convolving the recording with the complete binaural room impulse response (BRIR).

A pilot study to the current experiment was done in our listening room where binaural synthesis was done by implementing BRIRs. Head movements were allowed in both real life and binaural synthesis and the results were compared to fixed-head localisation. It was found that it is not necessary to update the whole BRIR according to the head movements. Instead it was sufficient to update only the direct sound component. The rest of the BRIR must, however, be correct with respect to the location of the loudspeaker in the room.

Therefore, every loudspeaker in the current experiment has its own measured reverberation tail and the direct sound is implemented by means of HRTFs measured in an anechoic chamber. The artificial head VALDEMAR (developed at Aalborg University) was employed for the measurements of the HRTFs and BRIRs.

The synthesis is therefore not made individually for every listener. However, the pilot experiment has shown that localisation errors are substantially reduced if head movements are allowed even though listeners' own heads (and ears) were not used to implement the synthesis.

2. PREVIOUS WORK

There are many accounts in the literature of localisation experiments with binaural recordings and synthesis, 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 are reported where it was possible for listeners to move their heads during binaural synthesis. However, Wenzel [3] demonstrated that enabling head tracking in binaural synthesis dramatically improves localisation even if nonindividual HRTFs are used. Sandvad [4] used individual 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 are possible.

Mackensen et al. [7] described a listening experiment where horizontal as well as vertical head movements were possible. The results indicate only a slight improvement when head tracking of both horizontal and vertical head movements is 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 individual 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 nineteen small loudspeakers were placed on stands around a chair in a standardised listening room. From a neighbouring control room a sound was played to one of the loudspeakers. The listener had to identify the loudspeaker from which the sound originated. Furthermore, the listening room was simulated using binaural synthesis. So the listener wearing headphones was also required to identify the loudspeakers when listening to this simulation.

Since it is the contribution of head movements that was of interest in this experiment listeners were sometimes allowed to move their heads and other times requested to sit still. This requires the binaural synthesis system to be able to react to head movements.

The task of identifying the loudspeakers had to be performed under the following five conditions:

1) real life, head move
2) real life, head still
3) binaural synthesis, head move, tracking on
4) binaural synthesis, head still, tracking on
5) binaural synthesis, head still, tracking off

In this way it is possible to compare real life with and without head movements. The same can be done for binaural synthesis. In conditions 3) and 4) the head tracking was enabled. This ensured in condition 4) that the sound was updated if small (unintended) head movements were made - even though the listener was requested to keep the head still.

In condition 5) the head tracking was not active and this condition therefore is similar to binaural recording. The results could therefore show the differences between real life and binaural synthesis and it should be able to determine whether small head movements (when the listener is requested not to move the head) have an influence.

3.1. Experimental set-up

The placement of the loudspeakers was the same as described in Møller et al. [1, 2] and Minnaar et al. [9]. Thirteen of the loudspeakers were located on the sphere around the listener's head at a distance of 1 m. 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. The set-up in the listening room can be seen in Figure 1.

Figure 1 Set-up of loudspeakers in the listening room. Listeners were allowed to move freely while seated on the chair.

The stimulus was a 5s long female speech sentence, which was recorded in an anechoic room. This signal was input directly to one of the loudspeakers in the set-up and played back at a natural level. Therefore, the loudspeakers were equalised with respect to their on-axis response (measured in an anechoic chamber).

A small 'traffic light' prompted the actions of the listener. After a sound was played an answer was submitted by pressing with a pen on an electronic tablet holding a schematic drawing of the loudspeaker set-up.

Under all conditions the listeners had to wear a head tracker and all head movements during the stimulus were recorded for later analysis. In conditions 3)-5) headphones were worn in addition to the head tracker. Furthermore, in conditions 3) and 4) the head tracking data was used as input to the binaural synthesis system in order to update the sound according to the head position.

The experiment was run from a control room adjacent to the listening room (see Figure 2). Two personal computers were used for controlling the experiment and implementing the binaural synthesis. The control software also interfaced a relay unit that channelled a signal to any one of the loudspeakers (for real life conditions) and it collected the answers of the listener. An intercom system was installed for communication between the listening room and the control room. Furthermore, two cameras were used to observe the actions of the listener.

Figure 2 Set-up of equipment in the control room adjacent to the listening room.

Twenty listeners (10 male and 10 female) participated. They were all between 20 and 30 years old and had controlled normal hearing.

Every listener went through a process of familiarization during which all the conditions were presented. Thereafter every listener was presented with 5 sessions (5 conditions) of approximately 12 minutes each. 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.

In conditions where it was not allowed to move the listener was simply requested sit still and to look straight ahead. When listeners were allowed to move they were encouraged to use any natural movements although they remained seated.

3.2. Implementation of binaural synthesis

The artificial head VALDEMAR was placed in the listener's position on the chair between the loudspeakers. A BRIR was measured for every loudspeaker, using maximum length sequences (MLS).

For this purpose a Windows-based MLS measurement system was developed that can measure two long impulse responses simultaneously [10]. It also ensures that the initial time delay in the impulse response is correct. It is important for the implementation of the synthesis that the absolute time of flight is correct for every loudspeaker.

Figure 3 shows a BRIR for the left ear, measured for a loudspeaker directly to the left of the artificial head. The data is plotted with its amplitude on a logarithmic scale in order to show the exponential decay. As described earlier the direct sound was removed from every BRIR - an example is shown in Figure 4. This later part of the impulse response (reverberation tail) was convolved with the inverse filter of the loudspeaker.

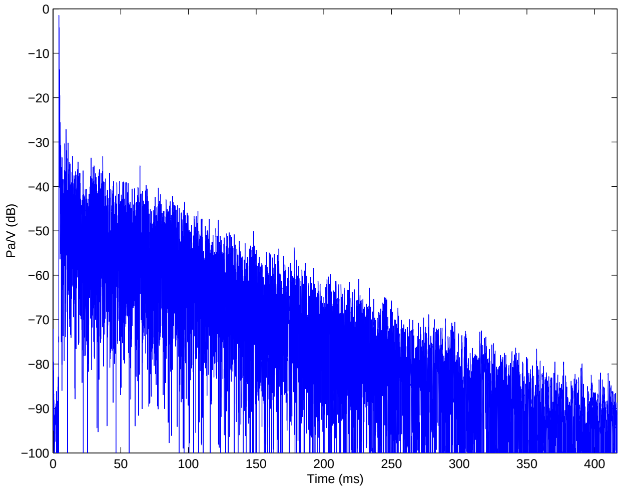

Figure 3 Measurement of a binaural room impulse response (BRIR) shown on a logarithmic scale.

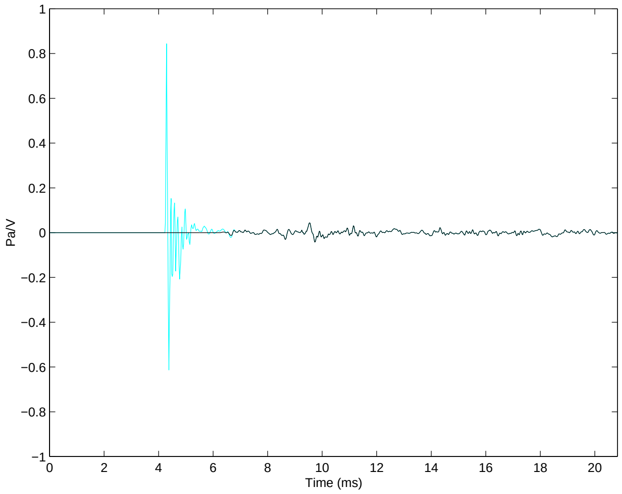

Figure 4 The BRIR in Figure 3 is shown here in light grey. The reverberation tail used during the experiment is shown in black.

Apart from measuring the BRIRs the HRTFs of the artificial head were measured in an anechoic room. Two-channel measurements were made for nearly 12000 directions on the sphere around the head [11]. The HRTFs were used during the binaural synthesis to implement the direct sound from each loudspeaker.

The complete BRIR was reconstructed by filtering the anechoically recorded signal separately with the direct sound and the reverberation tail. The two signals were then added and presented through headphones. The headphones were equalised from measurements made on the artificial head.

4. EXPECTED RESULTS

The results of all previous listening experiments where head movements were possible with binaural synthesis are very encouraging. Therefore this experiment employing 20 listeners was designed. It is now underway and the final results are expected to be available very soon. If the final results follow the trends seen in the pilot experiment they will show that localisation generally improves when head movements are allowed. This demonstrates the importance of dynamic auditory cues.

It is further expected that the improvement of binaural synthesis will be substantial; implying that much more realistic simulations can be made if the changes in the sounds at the ears due to head movements are implemented. The improvement due to head movements of distance perception may be less dramatic, however.

The pilot test showed that the number of localisation errors is relatively low in dynamic binaural synthesis even though listeners' own heads (and ears) were not used to implement the synthesis. The same result was found in previous studies [5, 6 and 7]. However, in those studies the complete BRIR was updated according to the head movements, whereas in this work only the direct sound has to be updated in real time, which leads to a substantially reduced computational load.

5. ACKNOWLEDGEMENTS

Sincere thanks are due to Jan Plogsties for his contribution to this work during the preparatory phase. Furthermore, we want to thank Claus Vestergaard Skipper for all the help in the laboratory.

6. 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] J. Sandvad, "Dynamic aspects of auditory virtual environments", *Presented at the 100th Convention of the Audio Engineering Society*, Copenhagen, Denmark, May 1996 (preprint 4226).

[5] 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).

[6] 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).

[7] 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).

[8] 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).

[9] P. J. Minnaar, S. K. Olesen, F. Christensen, H. Møller, "Localization with binaural recordings from artificial and human heads", *J. Audio Eng. Soc.*, May 2001.

[10] S. K. Olesen, J. Plogsties, P. J. Minnaar, F. Christensen, H. Møller, "An improved MLS measurement system for acquiring room impulse responses", *Presented at NORSIG2000*, Sweden, April 2000.

[11] B. P. Bovbjerg, F. Christensen, P. J. Minnaar, X. Chen, "Measuring the head-related transfer functions of an artificial head with a high directional resolution", *Presented at the 109th Convention of the Audio Engineering Society*, Los Angeles, California, USA, September 2000.