Distance perception in virtual auditory environments with a moving avatar

Annika Neidhardt
University of Technology Ilmenau, 98684 Ilmenau, Germany, Email: annika.neidhardt@tu-ilmenau.de

Introduction

Virtual auditory environments (VAEs) are of increasing interest, in science but also for industrial applications. Potential future buildings or products can be explored e.g. in virtual auditory walk-through scenarios.
The practical value of a VAE system is depending on its overall quality with criteria like plausibility or immersion. But additionally, it is of interest how well details of a virtual auditory scene can be perceived.

In the past it has been shown that head movements improve the localization of sound sources in real world scenarios as well as in VAEs with head tracking. Furthermore it has been shown, that in the real world the translation of a listener improves his auditory judgment of distance, e.g. by an effect called *acoustic tau.*

Even if a perfect reconstruction of the sound pressure at the ear entrance or ear drums could be achieved, in VAEs there might be drawbacks by a reduced scope of interaction or conflicting input of different modalities.
The question whether the effect of the acoustic tau is available in VAEs is examined in this paper. In-ear-recordings of moving people were evaluated in a listening experiment to find out if the translation of an avatar leads to an improved distance perception.

VAEs with a moving avatar

With VAEs, perceptions not corresponding to physical environment are created by acoustic representations [1]. The auditory scenes can be reproduced with a simple or spatially complex loudspeaker arrangement, but also via headphones using binaural technology.

The term *dynamic binaural synthesis* is used in a lot of studies, where actually only head rotations are taken into account. In contrast, the influence of a changing head position on the perception of VAEs has received little attention.
Considering a position-dynamic binaural synthesis, the user listens to the perspective of a moving avatar, as illustrated in fig. 1. Different kinds of interaction can be discriminated:

- Passive listening to a scene with a moving avatar - no interaction
- Non-authentic interaction - controlling the avatar with keyboard, joystick oder other input devices
- Authentic interaction - avatar is moving in correspondence to tracked movements of the listener, e.g. head tracking

Additionally, there are subtypes like combining head tracking with keyboard controlled translation. Authentic interaction also includes self-created sounds. The type of interaction is assumed to have an influence on the perception of VAEs as well. Therefore, such a discrimination is important.

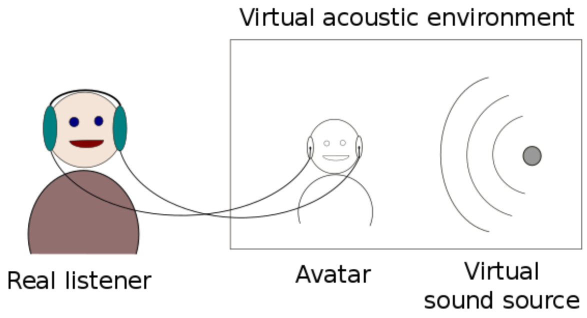

Figure 1: Binaural reproduction of a virtual auditory environment considering an avatar

Distance perception and dynamic cues

Zahorik [2] provides an overview of important parameters in the perceived distance of sound sources within an auditory display. Sound intensity, Direct-to-Reverberant Energy Ratio, the spectrum and binaural differences are stated as four relevant acoustic cues.
Additionally, some motion related information have been found to influence the human auditory distance perception.

Motion parallax effect

For a listener carrying out a translation not directly toward the sound source, the angular direction of the source changes. With an understanding of the own speed and direction of movement, this information can be used to estimate the egocentric distance to the source [3].
Furthermore, if there are more sound sources, the relative angles can change as well. In a reverberant room, strong reflections might also enable the listener to make use of the motion parallax effect.
The effect gets more complex, when moving sound sources are present. In this study only stationary sources are considered.

Acoustic tau

The acoustic tau specifies the time-of-arrival or time-to-contact of observer and source. Shaw et al. [4] derived the following equation:

$$\tau = \frac{2I}{dI/dt} \quad (1)$$

I is the time-averaged intensity at the observer. The equation considers free field conditions. Reflections, diffraction or scattering will result in a more complex relationship. But in case of weak reflections or a high direct-to-reverberant ratio, it is probably still a good approximation. The listener is assumed to move with a constant velocity towards a point source.
Because the intensity is square-inverse of the distance to the source, it increases rapidly at small times-to-contact. Thus, the acoustic tau might be easier to notice for a movement close to the source.

Previous studies

In different studies, e.g. [5, 6], head movements did not significantly enhance the human auditory distance perception. Considering VAEs for dynamic binaural reproduction, again no significant changes due to head movement could be found [7].

Ashmead et al. [8] showed that listening during a translational movement allows a more accurate distance perception than listening in one or two different static positions. There was only one sound source and the listeners were walking toward it. Therefore, it was concluded that the improved accuracy resulted from the acoustic tau.
Speigle and Loomis [3] conducted a similar experiment and found that, especially for the closer sound sources, participants walked slightly farther for the stationary condition. With a variation of the azimuth of the sound source the motion parallax effect was studied as well, but no significant impact could be found under the tested conditions.

In the experiment presented in [9] participants had to judge the distance to a virtual object in a sound field created using wave field synthesis. The probands spontaneously began to walk around for an exploration of the sound field. The authors concluded, that this highlights the importance of dynamic cues in virtual environments.

So far, dynamic effects due to a translation of the avatar have not been studied in VAEs for headphone reproduction. Additionally, the studies in [8] and [3] were conducted outside, nearly under free field conditions. Hence, experiences with respect to dynamic effects on auditory distance perception in rooms are required.
Rosenblum et al. observed a slightly increased accuracy in an echolocation task for moving compared to stationary listeners [10]. According to the authors, the time-to-contact might play some role.
Furthermore, motion parallax effects in matters of strong reflections could be supportive. It also has to be taken into account, that auditory distance perception in rooms underlies some training effect [11].

The Experiment

To get a better idea of dynamic effects in VAEs, the test items in a listening experiment should contain the acoustical details of a real movement in a real environment, since they might be important. Therefore, it was decided to work with binaural recordings.
One option was to use a dummy head. It is not easy to create a translation of the dummy head without disturbing sounds caused by the movement. With some effort there will be a solution to this problem. But still, the dummy head would not carry out the natural movements of a walking human being. There are slight changes in the angle relative to the sound source. Thus, for this pilot study it was chosen to use recordings captured by a walking person with in-ear-microphones.

Katz et al. [12] asserted, that blind people did not understand the auditory scene when they listened to the recordings of a moving person with in-ear-microphones. The option of own movements as well as the availability of dynamic cues relative to displacement seem to be essential for a comprehensive exploration.
In own pretests those observations were confirmed [13]. When listening to the in-ear-recordings of a person walking through the room, the probands were confused. Most often they did not understand, that they were listening to a changing perspective within the scene. Instead, most people thought that the sound sources were moving.

But, as described in the next paragraph, the scenes recorded for this experiment were kept very simple. The task to estimate the egocentric distance to the sound source was easily understood and carried out by everybody.

Binaural recordings

The binaural sound examples were created by recording the scenes with in-ear-electret-microphones at the entrance of the ear. One loudspeaker (Genelec 1030A) was positioned in a media lab with a volume of $V = 740m^3$ and a reverberation time of $T_{60} = 0,7s$. One short percussive sound was looped with a time interval of $750ms$ and played to the room via the loudspeaker. The sound was chosen to be least annoying.
The recording person was turned toward the loudspeaker and looking at it with a straight head at all times. To avoid any sounds of movement, the cables were fixed at the clothes and the person walked without shoes. Minimal head rotations occurred due to walking. The probands were not confused by that.
Furthermore, each distance and each movement was recorded three times. One version was used for the training, the other two were evaluated in the test. This was done to ensure, that the participants were trained to the distance and not to the recording itself.
In pretests is was observed that it was demanding and tiring to listen to a quick translation. Hence, the recordings were captured during slow movements, that were easier to follow. The person walked 2m in about five seconds.

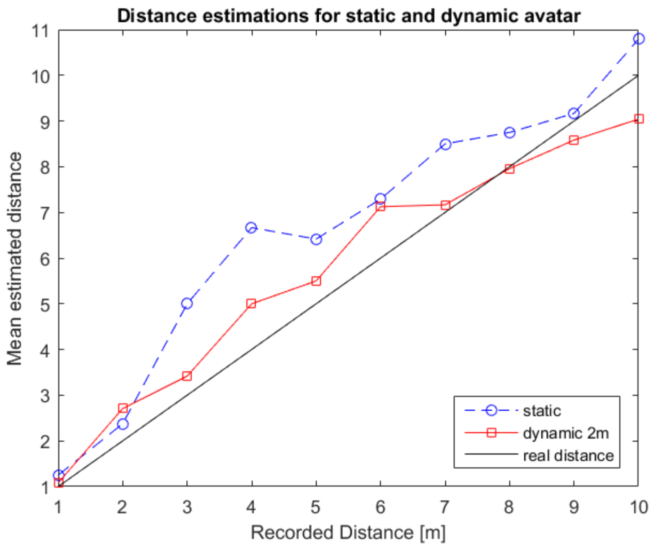

Figure 2: Distance perception while listening to a static and a dynamic avatar

The setup for the listening experiment

The experiment was carried out in speech recording lab (Rec. ITU-R BS.1116-1). For the binaural playback AKG K271 MII headphones were used. A graphical user interface (GUI) was programmed for this application.

The test items

For the stationary condition items with a length of seven seconds were recorded in distances of 1m to 12m away from the sound source with intervals of 1m. For the two dynamic conditions the sound pressure was captured while the recording person walked for 2m or 1m toward the loudspeaker. The egocentric distance was decreased e.g. from 8m to 6m during such an item. In that case, the participants were asked to judge the distance at the final position after the movement.

The test

15 people took part in the experiment, 11 male and 4 female. The average age was 26,7 years. Four stated, that they had no experience with listening tests or perception of binaural audio.
A single stimulus test design was chosen to avoid comparison. First, participants went through a training. It was divided into two separate training blocks. One of those blocks contained only samples recorded in a stationary position. The participants listened to 12 items with the distances of 2m, 4m, 6m, 8m, 10m and 12m. Six answer buttons were provided to choose the correct distances.
In the second training block recordings of the 2m translational movement towards the same distances were provided. The distance of the final position had to be estimated.
The actual test consisted of three separate blocks for the stationary recordings, the translation over 1m and 2m. The items within each block and the blocks themselves were shuffled in order for each participant.

Results

3 participants rated the distances with an average error greater than 2m, which was much higher than for the other participants. Those results were eliminated for the following analysis and plots.
Fig. 2 shows that the distance estimations after listening to a 2m-translation of the recording person were more accurate than the distances estimated for the static recording position.
An analysis of variance unveiled, that in the case of the stationary avatar there were significant differences to the real distance at 6 out of 10 distances. In comparison, in both dynamic conditions, significant differences to the real distance could be found only for 3 out of 10 distances.

It is also interesting to notice, that except for the 2m-distance, the estimated values for the static avatar are higher than for the 2m-dynamic condition.

	all	P1	P2	P3	P4
static	1.63m	1.33m	1.40m	1.20m	1.40m
dyn2m	1.22m	1.13m	0.80m	1.33m	1.33m
dyn1m	1.37m	0.73m	1.27m	1.27m	1.27m
over all	1.41m	1.06m	1.16m	1.27m	1.33m

Table 1: Mean error for the three test conditions - over all and for the four best individual results

Table 1 provides an overview of the average distance errors. For the static examples the average error over all participants of 1.64m is higher than the average error for the dynamic sound examples. The columns P1-P4 show the results of the best four participants. It is obvious, that they achieved their individual best performance in different conditions.
When comparing the results of the two dynamic conditions, for some distances the differences are significant. The recorded distance of 3m was estimated significantly more accurate for the longer movement, but e.g. for 6m the results were significantly more accurate for the 1m-movement. It can be seen in fig. 3 that no clear tendency for an influence of the length of the movement is observable. The average error over all distances (tab. 1) is minimally higher for the shorter translation.

Discussion and future work

The presented experiment adresses dynamic cues resulting from a translation of the avatar in a VAE. Participants listened passively to a binaural recording captured by a moving person wearing in-ear-microphones. Regarding interaction, that is a worst case scenario. Additionally, no individualization or headphone equalization was applied to the recordings.

Still, significant differences in the estimated distances for a stationary and a translating avatar could be found. Dynamic effects seem to influence the distance perception even in this simple realization of a VAE. Perhaps, the application of an individualization, a headphones equal-

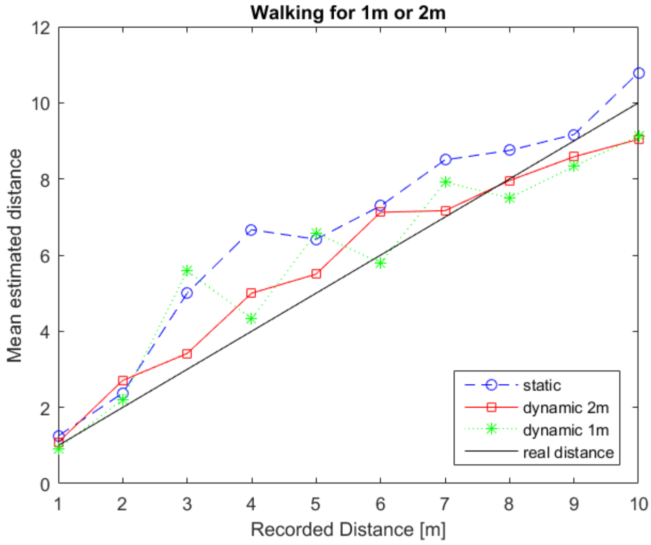

Figure 3: Distance perception while listening to a translation over two different distances

ization and a different training procedure would lead to even stronger differences.

For a creation of a VAE based on simulations, it is common to model the distance of a sound source by adjusting the level of the direct sound corresponding to an $1/r$ relation. It remains an open question, whether an influence of the listener movement on the perceived distance could still be observed in such a realization.
Although the average distance error is slightly smaller for the longer translation, there is no obvious tendency that the estimated distances after the shorter translation are less accurate. Hence, it would be interesting to know, if already small changes of the listening position improve the distance perception.
Furthermore, in this experiment, the translation of the avatar toward a sound source was addressed. Effects of a translation to the side or backwards should be studied as well. Additionally, the influence of dynamic aspects might change with the speed of movement.
In the case of an interactive scene exploration the avatar can be controlled by own tracked body movements or with input devices like a keyboard. Such an exploration could provoke an additional improvement.
Since this experiment was conducted with recordings captured in a room, reverberation might have an influence as well. A detailed analysis would be of interest.

The single stimulus test design was chosen for this experiment in order to avoid comparison. But for subsequent items, there might still be some interaction. Thus, a test design is required, that takes this aspect into account.
The experiment shows, that it is of interest to study the dynamic cues resulting from the translation of an avatar in-depth.

Acknowledgements

The authors would like to thank all the participants for their help.

References

[1] Pedro Novo. Auditory virtual environments. In Jens Blauert, editor, *Communication Acoustics*. 2005.

[2] Pavel Zahorik. Auditory display of sound source distance. In *Int. Conference on Auditory Distance, Kyoto, Japan, July*, 2002.

[3] Jon M. Speigle and Jack M. Loomis. Auditory distance perception by translating observers. *IEEE*, 1993.

[4] B.K. Shaw, R.S. McGowan, and M. Turvey. An acoustic variable specifying time-to-contact. *Ecological Psychology, Vol. 3, Issue 3, pp.253-261*, 1991.

[5] J. A. Simpson and L.D. Stanton. Head movement does not facilitate perception of the distance of a sound source. *American Journal of Psychology, 86, pp. 151-159*, 1973.

[6] L.D. Rosenblum, A.P. Wuestefeld, and K.L. Anderson. Auditory reachability: An affordance approach to the perception of sound source distance. *Echological Psychology, 8, pp. 1-24*, 1996.

[7] G. Kearney, X. Liu, A. Manns, and M. Gorzel. Auditory distance perception with static and dynamic binaural rendering. In *57th Int. AES Conf. on the Future of Audio Entertainment Technology – Cinema, Television and the Internet, March*, 2015.

[8] Daniel H. Ashmead, DeFord L. Davis, and Anna Northington. Contribution of Listeners' Approaching Motion to Auditory Perception. *Journal of Experimental Psychology, Human Perception and Performance, Vol. 21, No. 2, 239-256*, 1995.

[9] M. Rébillat, E. Corteel, and B.F.G. Katz. Audio, visual and audio-visual egocentric distance perception by moving subjects in virtual environments. *ACM Transactions on Applied Perception, October*, 2012.

[10] L. D. Rosenblum, M. S. Gordon, and L. Jarquin. Echolocating distance by moving and stationary listeners. *Ecological Psychology, 12 (3), 181-206*, 2000.

[11] Barbara Shinn-Cunningham. Learning reverberation: Considerations for spatial auditory displays. In *Int. Conference on Auditory Display (ICAD), Atlanta, Georgia, USA, April*, 2000.

[12] B.F.G. Katz and L. Picinali. Exploration of virtual acoustic room simulations by the visually impaired. In *Int. Seminar on Virtual Acoustics, Valencia, Spain, November*, 2011.

[13] Tahereh Afghah. Perception of distance after a continuous change of distance. Master's thesis, University of Technology Ilmenau, 2015.