Experiments on localization accuracy with non-individual and individual HRTFs comparing static and dynamic reproduction methods

Josefa Oberem[1], Jan-Gerrit Richter[1], Dorothea Setzer[1], Julia Seibold[2], Iring Koch[2] and Janina Fels[1]

[1] *Institut of Technical Acoustic, RWTH Aachen University, Germany, Email: job@akustik.rwth-aachen.de.*

[2] *Institute of Psychology, RWTH Aachen University, Germany.*

Introduction

The progress of individual binaural reproduction of recent years has brought the capability to study complex cognitive processes which require real-life conditions and high level of plausibility under a controlled and reproduce-able environment. The accuracy and plausibility of these binaural reproductions is highly dependent on several factors. These factors range from, for example, the set of the used head-related-transfer-functions (HRTFs) (e.g. individuality, resolution, section) as well as the type of reproduction (e.g. dynamic, static). The importance of each of these components individually has been discussed for several years. The benefit of the use of individual HRTFs for localization tasks has been discussed for several decades with some of the earliest work by Wightman, Kistler and Wenzel [1, 2]. A gain from small head movements was described first in Wallach [3] and could be reproduced several times [4, 5].

However, no study has systematically looked at the influence of the *resolution* of high resolution individual HRTFs and tried to quantify the influence of *both* high resolution individual HRTFs as well as dynamic binaural reproduction. Consequently, when dealing with a reproduction problem that requires real-life conditions, one must assume that all factors will benefit equally and are crucially important.

As taking all these factors into account is a time-consuming and expensive procedure the goal of this publication is to find a HRTF dataset which, combined with a reproduction method (static / dynamic), will result in a performance comparable to performance achieved with real sources. To this end localization experiments were conducted since localization performance delivers a rather general measure to indicate the quality of binaural auditory displays, not limited by any paradigm conditions. In the experiment individually measured HRTFs were compared to those of an artificial head in a static reproduction of stimuli and in three dynamic reproduction methods of different resolutions (5°, 2.5° and 1°).

Experimental design

The tested conditions were firstly differences in HRTF, looking into differences between individual vs. non-individual HRTFs. The reproduction method was the second condition. Here, differences between static reproduction, where head movements are ignored and dynamic reproduction, where head movements are compensated and the source will stay in a fixed position in the virtual space were investigated. For this condition, the HRTF resolution (5°, 2.5° and 1°) was also be analyzed. The influence of the source position, grouped into 3 categories (Front vs. Side vs. Back) was the last factor. For each condition and position 4 repetitions were done. The total number of tested stimuli can be calculated as 2 HRTFs $\times$ 4 reproduction methods $\times$ 11 positions $\times$ 4 repetitions + 48 training trials = 400 trials.

Participants were asked to localize a train of pulsed white Gaussian noise. The frequency domain ranged from 100 Hz to 20 kHz providing high-frequency pinna cues essential for localization [6]. The use of white noise was in accordance with previous studies and strongly recommended for the investigations of localization with head movements as broadband signals will excite all changes in monaural in binaural information during tiny head motion (e.g. [7]). The stimuli were convoled with respective HRTFs and auralized using a real-time auralizer developed at the Institute of Technical Acoustics, RWTH Aachen University. They are played back with headphones, individually equalized to the subject [8]. As head movements were found best observable for minimum duration of 2 s [9], the total stimulus length was set to 3.7 s and was composed of five alternating 0.3 s- and 1.2 s-bursts of constant sound level with on- and offset ramps (50 ms rise and 50 ms fall time each), interrupted by small pauses of 100 ms.

Proximal pointing was used as a pointing method. The egocentric method was introduced by Bahu and colleagues [10] and is recommended for closed-loop localization tasks. The listener indicates spatial positions by placing a hand-held marker in the region of the head where (s)he perceives the incident sound.

The source positions were limited to the right hemisphere because of limitations of the pointing method. In total, 11 positions were tested which can be grouped in 3 categories (Front vs. Side vs. Back). There were four frontal positions, two on the median plane with difference in elevation of 30° as well as two positioned close to the median plane ($\varphi = 350°$). Sources in back were designed to be symmetrically identical with those in front. However, the source with the lower elevation ($\vartheta = 120°$) was not added due to limitations of the pointing method. Sources to the side were also designed to be symmetrically identical and to be positioned on one cone of confusion. 11 right-handed, student participants aged between 20 and 27 years with an average of (23.6

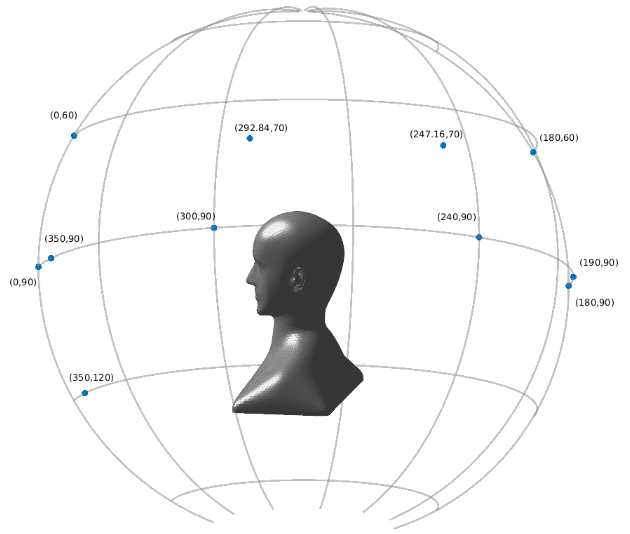

Figure 1: Source positions around listener in localization experiment.

± 2.4 years) completed the localization experiment. Participants were equally divided in male and female listeners and the experiment was remunerated with 40 Euro. All participants were inexperienced in hearing with HRTF-data and not trained in localization prior to the experiment. Hearing sensitivity (< 15 dB HL) has been verified with the aid of a pure tone audiometry between 125 Hz and 10 kHz. For the test, individual HRTFs [11] are measured for each subject and for the used artificial head. The HRTFs are measured with 2.5° resolution in approximately 10 minutes. For higher resolution, an interpolation using spherical harmonics is used [12].

Results

This section will present the localization error for unsigned horizontal error, unsigned vertical error, weighted front-back-confusions and in-head localization metrics.

Unsigned horizontal error

Figure 2 shows unsigned horizontal error as a function of the position, reproduction method and the HRTF.

Although there was a non-significant trend towards smaller horizontal errors when listening to individual HRTFs (individual HRTFs: 24.2 ° vs. non-individual HRTFs: 27.4 °), there was no significant effect on HRTFs in the unsigned horizontal error.
The main effect of reproduction method was significant, resulting in a significantly higher unsigned horizontal error for a static reproduction than for all three dynamic reproduction methods (static: 38.7 ° vs. dynamic (5°): 21.2 ° vs. dynamic (2.5°): 21.5 ° vs. dynamic (1°): 21.7 °). The interaction of HRTF and reproduction method was significant. Post-hoc test showed that the unsigned horizontal error difference between individual and non-individual HRTFs was significantly greater for the static reproduction than for the dynamic reproductions (static: 9.86 ° vs. dynamic (5°): 0.20 ° vs. dynamic (2.5°): 0.64 ° vs. dynamic (1°): 2.18 °).

Table 1: F and p-Values for unsigned horizontal error analysis

HRTF:	$F(1,10) = 2.50$,	$p > .05$
RepMeth:	$F(3,30) = 15.85$,	$p < .001$
HRTF × RepMeth:	$F(3,30) = 3.70$,	$p < .05$
Pos:	$F(2,20) = 2.57$,	$p > .05$
HRTF × Pos:	$F < 1$	
RepMeth × Pos:	$F(6,60) = 3.48$,	$p < .05$
HRTF × RepMeth × Pos:	$F(6,60) = 2.18$,	$p > .05$

The main effect of source position was not significant. However, unsigned horizontal errors were smallest for trials where the source was positioned to the side and greatest for trials where the source was positioned in front (Front: 36.8 ° vs. Side: 16.9 ° vs. Back: 23.6 °). As the error was not corrected for front-back confusions this result is not surprising. The interaction of HRTF and source position did not turn out to be significant, meaning that there was gain from the use of individual HRTFs at specific locations only. The interaction of reproduction method and source position was significant, indicating no significant differences between reproduction methods for source positions to the side, but several significant differences between the static and dynamic reproduction for source positions in front and in back. This is explained by a reduction of the number of front-back confusions with dynamic playback. All p and F-Values for the analysis can be found in Table 1.

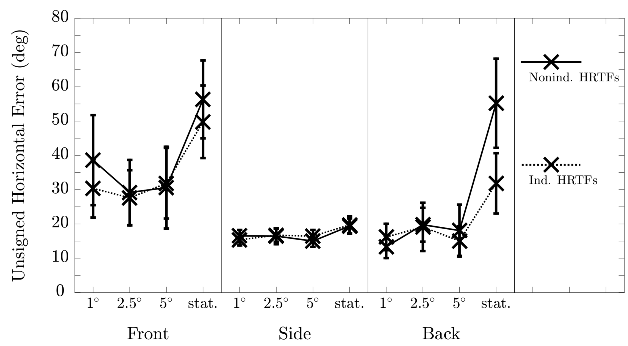

Figure 2: Unsigned horizontal error (in deg) as a function of HRTF, reproduction method and source position (**HRTF × RepMeth × Pos**). Error bars indicate standard errors.

Unsigned vertical error

Figure 3 shows unsigned vertical error as a function of the position, reproduction method and the HRTF. There was no significant effect in the unsigned vertical error for either the used HRTF, the reproduction method or the interaction between both. This indicates, that there was no gain in the elevation localization accuracy with the use of individual HRTFs regardless of reproduction method. The main effect of source position was significant, indicating significantly smaller unsigned vertical errors for trials where the source was positioned to the side than for trials where the source was positioned in front or in back (Front: 19.2 ° vs. Side: 10.3 ° vs. Back: 20.6 °). All p and F-Values for the analysis can be found in Table 2.

Table 2: F and p-Values for unsigned vertical error analysis

HRTF:	$F < 1$	
RepMeth:	$F < 1$	
HRTF × RepMeth:	$F(3, 30) = 2.03,$	$p > .05$
Pos:	$F(2, 20) = 13.65,$	$p < .001$
HRTF × Pos:	$F < 1$	
RepMeth × Pos:	$F < 1$	
HRTF × RepMeth × Pos:	$F(6, 60) = 1.72,$	$p > .05$

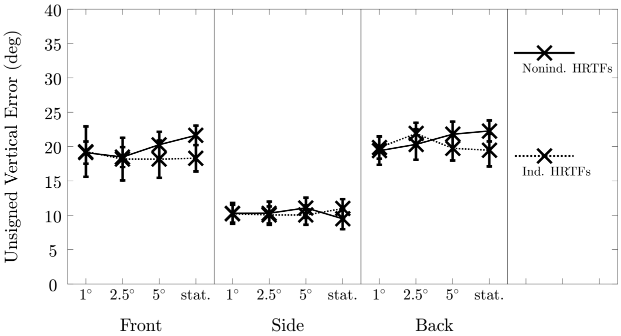

Figure 3: Unsigned vertical error (in deg) as a function of HRTF, reproduction method and source position (**HRTF** × **RepMeth** × **Pos**). Error bars indicate standard errors.

Weighted front-back-confusions

Figure 4 shows weighted front-back confusion in percent as a function of the position, reproduction method and the HRTF. The arc-sine transformation of data which is appropriate for the data on proportions is applied before conducting the ANOVA. The ANOVA yielded a significant effect on HRTFs in the weighted front-back-confusions, indicating less front-back-confusions when listening to individual HRTFs (individual HRTFs: 7.6 % vs. non-individual HRTFs: 10.7 %).

The main effect of reproduction method was also significant, resulting in a significantly higher front-back-confusion for a static reproduction than for all three dynamic reproduction methods (static: 16.5 % vs. dynamic (5°): 6.3 % vs. dynamic (2.5°): 6.8 % vs. dynamic (1°): 7.1 %). The interaction of HRTF and reproduction method was not significant. However, there was a non-significant trend towards a greater front-back-confusion difference between individual and non-individual HRTFs for the static reproduction than for the dynamic reproductions (static: 6.8 % vs. dynamic (5°): 1.7 % vs. dynamic (2.5°): 2.0 % vs. dynamic (1°): 1.6 %). This leads to the interpretation that dynamic reproduction will reduce the gain from individual HRTFs in terms of front-back confusion.

The main effect of source position was not significant. However, weighted front-back-confusions were smallest for trials where the source was positioned to the side and greatest for trials where the source was positioned in front (Front: 14.5 % vs. Side: 4.7 % vs. Back: 8.3 %). The two-way interaction of reproduction method and source position is significant, indicating higher front-

Table 3: F and p-Values for weighted front-back-confusions

HRTF:	$F(1, 10) = 6.98,$	$p < .05$
RepMeth:	$F(3, 30) = 11.68,$	$p < .001$
HRTF × RepMeth:	$F(3, 30) = 1.31,$	$p > .05$
Pos:	$F(2, 20) = 1.31,$	$p > .05$
HRTF × Pos:	$F(2, 20) = 1.02,$	$p > .05$
RepMeth × Pos:	$F(6, 60) = 3.54,$	$p < .05$
HRTF × RepMeth × Pos:	$F(6, 60) = 2.17,$	$p > .05$

back-confusions for a static reproduction in front and in back compared to the side. All p and F-Values be found in Table 3.

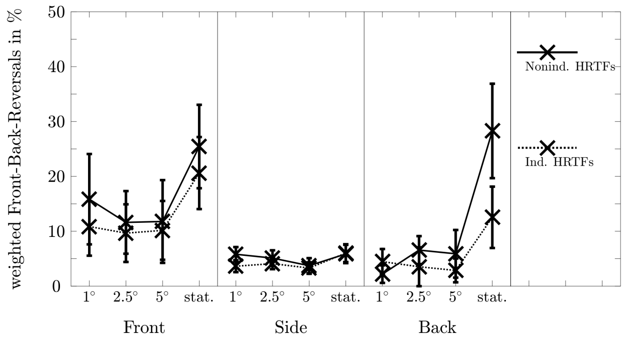

Figure 4: Weighted front-back-confusions (in %) as a function of HRTF, reproduction method and source position (**HRTF** × **RepMeth** × **Pos**). Error bars indicate standard errors.

In-head-localization

Figure 4 shows in-head localization occurance in percent as a function of the position, reproduction method and the HRTF. The arc-sine transformation of data which is appropriate for the data on proportions is applied before conducting the ANOVA. There was no significant effect on HRTFs in in-head-localization. However, there was a non-significant trend towards a lower rate of in-head-localization when listening to individual HRTFs (individual HRTFs: 1.5 % vs. non-individual HRTFs: 2.6 %).

The main effect of reproduction method was significant, resulting in a greater rate of in-head-localization for the static reproduction than for all three dynamic reproduction methods (static: 6.4 % vs. dynamic (5°): 0.5 % vs. dynamic (2.5°): 0.6 % vs. dynamic (1°): 0.7 %). The interaction of HRTF and reproduction method was not significant. However, there was a non-significant trend towards a greater in-head-localization difference between individual and non-individual HRTFs for the static reproduction than for the dynamic reproductions (static: 3.6 % vs. dynamic (5°): 0.5 % vs. dynamic (2.5°): 0.3 % vs. dynamic (1°): -0.2 %).

The main effect of source position was significant, indicating a significantly lower rate of in-head-localization for trials where the source was positioned to the side than for trials where the source was positioned in front (Front: 3.4 % vs. Side: 0.1 % vs. Back: 2.7 %). The two-way interaction of HRTF and source position did not turn out

Table 4: F and p-Values for in head localization

HRTF:	$F(1,10) = 1.92$,	$p > .05$
RepMeth:	$F(3,30) = 8.12$,	$p < .05$
HRTF × RepMeth:	$F(3,30) = 1.68$,	$p > .05$
Pos:	$F(2,20) = 6.59$,	$p < .05$
HRTF × Pos:	$F < 1$	
RepMeth × Pos:	$F(6,60) = 5.36$,	$p < .05$
HRTF × RepMeth × Pos:	$F(6,60) = 1.63$,	$p > .05$

to be significant. The interaction of reproduction method and source position was significant, indicating more in-head-localization with the static reproduction for sources positioned in front and back than to the side.
All p and F-Values for the analysis can be found in Table 4.

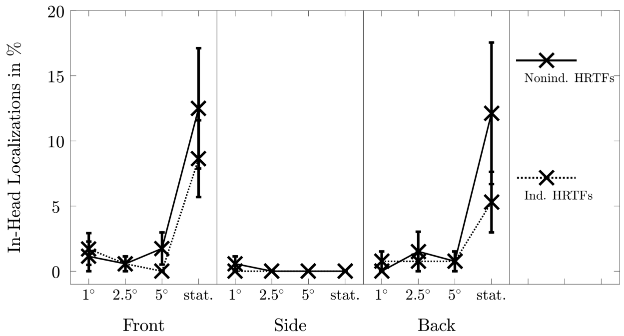

Figure 5: In-head-localization (in %) as a function of HRTF, reproduction method and source position (**HRTF × RepMeth × Pos**). Error bars indicate standard errors.

Conclusion

Aim of this investigation was to find an appropriate setting for the use of individual/non-individual HRTF and static/dynamic reproduction methods to ensure a most realistic auralization of acoustic scenes with binaural technique and headphone reproduction for further application in listening experiments.
Dynamic reproduction of any resolution applied was confirmed fundamental for a reduction of undesired front-back-confusions and in-head-localization. With front-back-confusions dominating the results for the unsigned horizontal error, dynamic reproduction implicitly also had a great impact on the azimuth error. Individually measured HRTFs showed a smaller effect on localization accuracy compared to the influence of dynamic sound reproduction. They were mainly observed to reduce the front-back-confusion rate. In general, individual HRTFs proved indeed beneficial compared with non-individual HRTFs, however, as the benefit tended to be masked in the presence of the more pronounced effect of reproduction method, using non-individual HRTFs in a dynamic reproduction might be a less complex and elaborate way to successfully provide for a realistic auditory perception. The resolution of the used HRTFs did not produce a significant increase in localization performance, however audible artifacts with a resolution of (5 °) were present indicating the need for a higher resolution.

References

[1] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 868–878, 1989.

[2] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, pp. 111–123, jul 1993.

[3] H. Wallach, "The Role of Head Movements and Vestibular Aad Visual Cues in Sound Localization," *J. Exp. Psychol.*, vol. 27, no. 4, 1940.

[4] I. Pollack and M. Rose, "Effect of head movement on the localization of sounds in the equatorial plane,"

[5] D. R. Begault, E. M. Wenzel, and 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.," *J. Audio Eng. Soc.*, vol. 49, pp. 904–916, oct 2001.

[6] R. B. King and S. R. Oldfield, "The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays," *Human Factors*, vol. 39, no. 2, pp. 287–295, 1997.

[7] F. L. Wightman and D. J. Kistler, "Resolution of front–back ambiguity in spatial hearing by listener and source movement," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2841–2853, 1999.

[8] B. Masiero and J. Fels, "Perceptually robust headphone equalization for binaural reproduction," in *130th Audio Engineering Society Convention*, (New York, NY, US), p. 8388, 2011.

[9] W. Thurlow and J. Mergener, "Effect of stimulus duration on localization of direction of noise stimuli," *Journal of Speech and Hearing Research*, vol. 13, pp. 826–838, 1970.

[10] H. Bahu, T. Carpentier, M. Noisternig, and O. Warusfel, "Comparison of different egocentric pointing methods for 3d sound localization experiments," *Acta Acustica united with Acustica*, vol. 102, no. 1, pp. 107–118, 2016.

[11] J.-G. Richter, G. Behler, and J. Fels, "Evaluation of a fast hrtf measurement system," in *140th Audio Engineering Society Convention*, (New York, NY, US), p. 9498, 2016.

[12] R. Duraiswaini, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of hrtfs [head related transfer functions]," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. iv–45–iv–48 vol.4, 2004.