Comparison of the front-back confusion rate of individual and individualized head-related transfer functions

Janina Fels, Hark Braren, and Ramona Bomhardt
Institute of Technical Acoustics, Teaching and Research Area for Medical Acoustics, RWTH Aachen University, Germany.
Email: {jfe, hbr, rbo}@akustik.rwth-aachen.de

Abstract

The individualization of head-related transfer functions (HRTFs) is worthwhile if no individual data is available. One option is to individualize an HRTF data set using anthropometric measurements of the person to be adapted and principal components (PCs). However, the localization performance using an individualized data set could be worse than for an individual one. This is often based on the fact that a mismatch between the individual and individualized HRTFs is still remaining. This issue is investigated by the confusion rate (front-back and back-front) using individual and various stages of individualized HRTFs in the following study described. The individualization approach for the HRTFs is based on an interaural time difference (ITD) model for low frequencies and the PCs in combination with anthropometric dimensions. Hence, not only the individual and individualized HRTFs are tested but also the influence of the ITD model and estimated HRTFs using optimally weighted PCs. Furthermore, to reduce the measurements for the anthropometric dimensions, the symmetry of HRTFs is investigated.

Localization using head-related transfer functions

Individual HRTFs provide a better localization performance and lower front-back confusion rates than non-individual ones [7]. However, the technical measurement effort is excessively high. In general, the front-back and back-front confusion rate depends on: head movements [6], direction of the incident sound [7], playback method [3], and listener's experience with the task [7]. Concluding, the least confusions occur under free-field conditions with free head movements.

HRTF modifications

Five different HRTF data set types were tested in the listening experiment.

1. The individual reference data sets are taken from ITA HRTF database [1]. In total, 48 individual HRTF data set with a spatial resolution of 5° are considered. Each transfer function has 256 samples using a sampling frequency of 44.1 kHz.
2. The influence on the symmetry of an HRTF data set was investigated mirroring the left ear HRTFs from the database [1] at the median plane so that both ears are symmetric to show the effect of the asymmetry of the head.
3. The ITD can be modeled by a minimum phase component plus delay. The delay is estimated by an ellipsoidal model using the head width, depth, and height [2].
4. The database is used to calculate the PCs and their corresponding optimal weights. For this study, 15 PCs are used to estimate an HRTF data set [2].
5. The PC weights are estimated using nine head and ear dimensions (dimensions of the cavum concha, cymba, head width and depth) [2]. According to the previous setup, the anthropometric HRTFs are estimated using 15 PCs.

Listening experiment

In total, 18 experienced subjects were tested with an average age of 30 ± 5 years. One of the subjects was female and all subjects' HRTFs are part of in the ITA HRTF database [1]. For the playback, a pink noise train stimulus was used with two pulses (0.3 s each) and a pause (0.3 s) in between. This stimulus was multiplied in the frequency domain with the subjects' individual headphone transfer functions. The transfer functions were measured using Sennheiser HD 650 headphones and Sennheiser KE3 microphones at the ear canal entrance according to Masiero and Fels [4]. Finally, the signals were convolved with an HRTF pair corresponding to one of tested directions. To cover sensitive source locations [5], directions in the horizontal plane at the subject's right side (10°, 30°, 70°, 110°, 150°, and 170°)[1] were chosen. The experiment took place in a low reverberant chamber using a static reproduction without head movements via headphones (Sennheiser HD 650). Every direction was tested 15 times per condition (direction and HRTF type). Before the experiment, a short training sequence was conducted. Here the subject's task was the same as in the actual experiment to distinguish between front and back and to indicate in-head localizations. To avoid training effects with the individual or modified HRTFs, the training was performed using real loudspeakers located in the same locations as the virtual sound sources used in the experiment. The average total duration of the complete test was 30 minutes.

Front-back and back-front confusions

The confusion rates, either front-back or back-front, for the five different HRTF types are calculated in percent

[1] In the ITA HRTF database would have a negative sign.

per subject and direction. While in Fig. 1 the average per direction over all subjects is evaluated, Fig. 2 shows the frequency of occurrence of the percentage confusion rates.

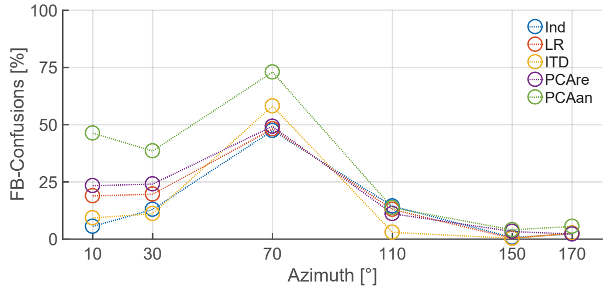

Figure 1: The mean confusion rate of the listening experiment is shown per HRTF type by circles. The individual HRTF configuration (Ind) is shown in blue, the symmetric one (LR) in red, the one with the ITD model in yellow, the estimated one (PCAre) in purple, and the anthropometric one (PCAan) in green.

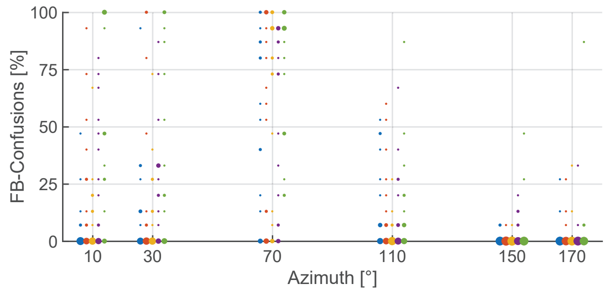

Figure 2: The percentage confusions rate per subject is shown by filled dots. The larger the dot, the more subjects had the same percentage. The individual HRTF configuration (Ind) is shown in blue, the symmetric one (LR) in red, the one with the ITD model in yellow, the estimated one (PCAre) in purple, and the anthropometric one (PCAan) in green.

Frontal positions

The confusion rate for individual HRTFs is comparable to the one using an ITD model for frontal positions. Both averaged rates are below 15%. The rate increases to 19% when the HRTF is symmetric. Noticeably higher is the rate for HRTFs estimated from 15 optimally weighted PCs compared to the previously mentioned ones. In this case, the mean rate amounts to 23%. The highest confusion rates are observed for the anthropometric HRTFs. For this type of HRTFs, 46% of frontal positions are rated as rear ones.

Lateral positions

Frontal lateral positions are often rated as rear ones. The results using an ITD model shows a tendency that the subjects rather shift lateral positions in the frontal hemisphere to the back than rear ones to the front. The hypothesis is that the subjects see their interaural axis located slightly more forward with regards to the HRTF coordinate system. This effect can be monitored as well for all other tested types of HRTFs at 70°.

Rear positions

Back-front confusions are rare and slightly lower at 150° than 170°. Except for the anthropometric HRTFs, those rates are below 6%. Consequently, for most of the subjects the number of back-front confusions is often 0%.

In-head localizations

According to the confusion rate, the in-head localization rate per subject is calculated as the percentage average of the number of in-head localizations. Subsequently, in Fig. 3 the average is determined from this rate per direction.
In general, the rate of in-head localizations is higher in the frontal hemisphere. At 10° the rate varies between 36 and 39% expect for the anthropometric HRTF which is approximately 50%. At the side, the rates are lowest for individual HRTFs (1%) and increase for anthropometric HRTFs to 11%. The number of in-head localization increases slightly for almost all conditions from 150° to 170°. The increase is again strongest for the anthropometric HRTFs.

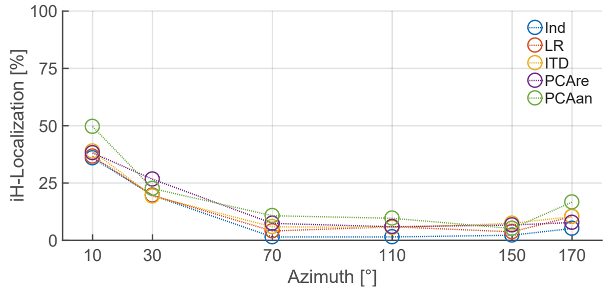

Figure 3: The mean in-head localization rate of the listening experiment is shown per HRTF type by circles. The individual HRTF configuration (Ind) is shown in blue, the symmetric one (LR) in red, the one with the ITD model in yellow, the estimated one (PCAre) in purple, and the anthropometric one (PCAan) in green.

Summary and conclusion

Confusions

Performance using individual or modeled ITD is comparable at almost all tested directions. The symmetric HRTFs show an increased number of front-back confusions. Consequently, the influence of the head's asymmetry is not negligible for some subjects. Especially for frontal directions, 15 PCs are not sufficient for optimal HRTF estimation. Using anthropometric HRTFs on the basis of PCs show considerably increased front-back confusion rates due to the additional loss of accuracy in the estimation step. Lateral directions close to the interaural axis seem to be challenging for all subjects due to a mismatch between their individual interaural axis and the one defined by the HRTF coordinate system. For these directions, the rate are highest. Back-front confusions

are rare for all tested HRTF types and positions.
The comparison of studies [3, 7] is difficult since the rate depends on the tested directions, experience and playback method.

In-head localizations

The in-head localization rate shows the ability to externalize the virtual sound source. Without externalization, the localization of sources in the front or back is challenging. Most of the in-head localizations occurred in the front which indicates a relation to the confusion rates.

References

[1] Bomhardt, R., de la Fuente Klein, M., Fels, J. (2016, November). A high-resolution head-related transfer function and three-dimensional ear model data-base. In Proceedings of Meetings on Acoustics 172 (Vol. 29, No. 1).

[2] Bomhardt, R. (2017). Anthropometric Individualization of Head-Related Transfer Functions Analysis and Modeling (Vol. 28). Logos Verlag Berlin GmbH.

[3] Kistler, D. J., Wightman, F. L. (1992). A model of head-related transfer functions based on principal components analysis and minimum phase reconstruction. The Journal of the Acoustical Society of America, 91(3), pp. 1637-1647.

[4] Masiero, B., Fels, J. (2011). Perceptually robust headphone equalization for binaural reproduction. In Audio Engineering Society Convention 130. Audio Engineering Society.

[5] McFadden, D., Pasanen, E. G. (1976). Lateralization at high frequencies based on interaural time differences. The Journal of the Acoustical Society of America, 59(3), pp. 634-639.

[6] Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. Journal of Experimental Psychology, 27(4), p. 339.

[7] Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. The Journal of the Acoustical Society of America, 94(1), pp. 111-123.