HRTF adaptation and pattern learning

FLORIAN KLEIN* AND STEPHAN WERNER
Electronic Media Technology Lab, Institute for Media Technology, Technische Universität Ilmenau, D-98693 Ilmenau, Germany

The human ability of spatial hearing is based on the anthropometric characteristics of the pinnae, head, and torso. These characteristics are changing slowly over the years and therefore it is obvious that the hearing system must be adaptable to some degree. Researchers have already been able to measure this effect, but still there are many open questions like the influence of training time and stimuli, level of immersion, type of feedback, and inter-subject variances. With HRTF (head-related-transfer-function) adaptation it might also be possible to increase the plausibility of acoustical scenes over time. When measuring adaptation effects in a spatial hearing test it is important to distinguish between conscious pattern learning and perceptive adaptation. To increase the quality of virtual auditory display the amount of perceptive adaptation is of major interest. In an earlier spatial listening test high training effects could be observed within a short period of training. To investigate the different types of training a second listening test was conducted. The acoustic stimuli were altered between the test and training sessions to avoid pattern learning. The results are compared to the previous findings and give further insights into the topic of perceptive adaptation of HRTFs.

MOTIVATION AND STATE OF THE ART

In auditory research adapation effects of the auditory system are well known for example in the field of cochlear-implant (CI) treatment. In other research areas like in the development of spatial sound systems adaptation effects are mostly not evaluated. In recent publications the existence of spatial-hearing adaptation effects could be observed by Majdak (2012) and Parseihian and Katz (2012), and earlier by Hofman *et al.* (1998). The focus of research is the localisation performance of the listeners and the achievable accuracy gain by listening training. Researchers found better performances for quadrant errors (e.g., front-back confusions) and elevation perception after audio-visual or proprioceptive feedback. Training in virtual environments like in Majdak (2012) and Parseihian and Katz (2012) exhibit fast training effects after a short time of training. Under real conditions, by using ear molds to modify the head-related transfer functions instead of binaural synthesis, training effects are observed after many days of training.

Listening tests in our lab (referring to Klein and Werner (2013)) confirmed significant

*Corresponding author: florian.klein@tu-ilmenau.de

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

adaptation effects regarding the perception of elevation after audio-visual training in a virtual environment. Strong adaptation effects after short training periods are often understood as pattern learning (or known as procedural learning in Hawkey *et al.* (2004)) in contrast to perceptual adaptation because of the training on a specific task. A second listening test is compared to a previous adaptation test with the following questions:

1. Are adaptation effects persistent after a long time without training and could that be and indication for perceptual adaptation?

2. Can accuracy gains be achieved when test and training stimuli differ?

TEST ENVIRONMENT

Static binaural synthesis is used to create the spatialisation of different directions. A block diagram of the rendering system is shown in Fig. 1. Artificial HRTFs (CIPIC database by Algazi *et al.* (2001) and own KEMAR measurements) are used in combination with a least-squares-based headphone equalization according to Schärer (2008). For all tests STAX lambda pro headphones are used.

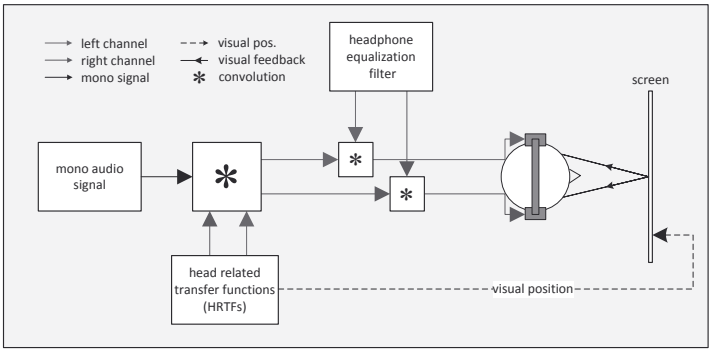

Fig. 1: Block diagram of the used binaural synthesis system combined with visual feedback over a screen; non-individual head-related transfer functions from the CIPIC database by Algazi *et al.* (2001) and headphone equalization according to Schärer (2008) are used.

For the audio-visual training participants are placed in front of a screen with loudspeaker symbols in direction of the virtual sound sources. Loudspeaker symbols for the virtually active loudspeaker are highlighted in green during the training sessions. A picture of a typical test situation is shown in Fig. 2. During the test sessions, the participants can simply use a computer mouse to select the loudspeaker

symbol which is nearest to the perceived direction. Visual loudspeaker representations are placed at azimuth angles of -30° to 30° in steps of 5° and at the vertical angles between 28.125° to -16.875° in steps of 5.625° .

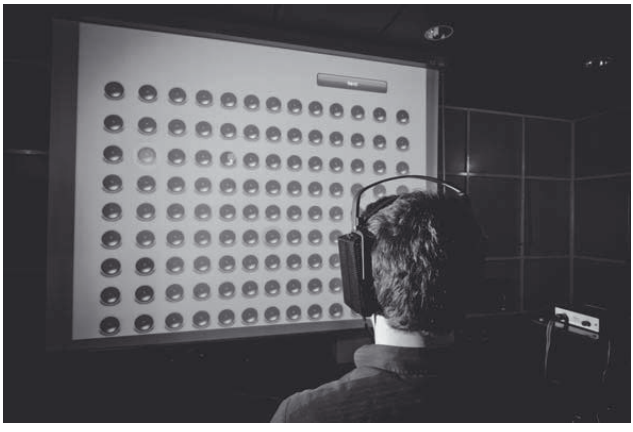

Fig. 2: Picture of the actual listening test setup in the listening lab of the TU Ilmenau.

Because a static binaural system is used and visual feedback is provided over screen, the positioning of the participants is crucial. Before each test the listeners are positioned at a defined distance and height in front of the screen. During the test the participant has to point his head towards the central loudspeaker symbol which is highlighted in red.

LISTENING TEST DESIGN

When using artificial HRTFs the localisation performance of the participants varies highly. Therefore initial pre-tests are conducted to measure the individual performance without any training and for each test stimulus. After the training sessions post-tests are conducted to measure the change in localisation performance. The different types of test and training sessions are described below.

Audio-visual training

For training, a sequence of virtual auditory sound sources is synthesized together with the spatially corresponding visual feedback. In one session 72 trials are presented which consist of four random azimuth directions for all nine elevation angles. Each direction gets repeated once.

Listening test with further audiovisual training

The audio stimuli are presented and the listener has to choose a perceived direction. After the response is given, the correct loudspeaker is highlighted in green. This session consists of 72 trials (six random azimuth angles at six different elevation angles and one repetition). In the test session virtual acoustic loudspeakers are synthesized only at the vertical angles of $28.125°, 16.875°, 5.625°, 0°, -5.625°$, and $-16.875°$. This kind of test can be understood as active training in contrast to the passive training by just watching and listening to a sequence of trials. Furthermore, this way test data can be acquired in the training phase.

Listening test without visual cues

During this test the participant has to rate 72 sound stimuli (equal to 'Listening test with further audiovisual training') and gets no visual feedback at all.

An overview of all test comparisons is shown in Fig. 3. The first test was done five months before the second test. The second test is divided into two parts to keep the test duration under 60 minutes. Overall 14 participants took part in the second test while nine of them already took part in the first test. For these nine listeners a comparison between test one and two can be done. The second test is aimed to compare the effect of different testing and training conditions by using different acoustic stimuli.

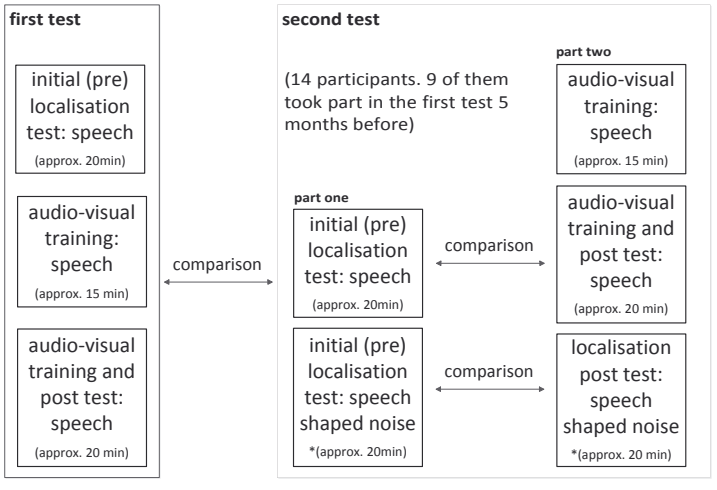

Fig. 3: Overview of the test procedures of the first and second test. The marked comparisons between tests are discussed in the results section.

The speech signal is about three seconds long and features a male foreign speaker. The other stimulus is CCITT coloured noise (according to ITU-T Rec. G227) with the

same temporal envelope as the speech signal. The power spectrum of CCITT coloured noise is similar to the average power spectrum of typical speech.

RESULTS AND EVALUATION

Because the test setup only allows ratings in the frontal plane, training effects on front-back confusions can not be evaluated (in Majdak (2012) results about the reduction of quadrant angle errors like front-back confusion can be found). Therefore the focus of this publication is directed at changes in the perception of elevation.

Comparison between first and second test

In the first comparison the ability to discriminate different elevation angles is investigated. Elevation angles are ranked according to their perceived height and compared to the correct order of the elevation angles. If a participant orders all elevation angles correctly according to their height (for example -16.875° is perceived at -11.25°, 0° is perceived at -5.625°, and 16.875° is perceived at 5.625°), then the rank correlation equals one. Perceiving different target angles at the same angle or alternating the order of elevations results in a lower rank correlation. This approach gives no statement about the absolute height accuracy and is therefore more liberal than localisation error scores. Figure 4 shows box plots of the rank correlations for the different tests conducted with the nine participants from the first and second test.

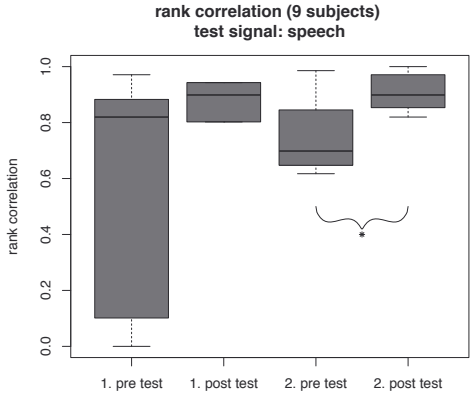

Fig. 4: Boxplot of rank correlations for each test. A rank correlation of one means that all participants were able to order the sound samples according to their height (no statement about absolute height accuracy). Significant differences are marked with an asterix symbol (Wilcoxon signed rank test with confidence interval < 0.05).

The comparison with the first test shows that the ability to discriminate elevation angles decreases without further training. There is no significant difference in rank correlation between the results of the first and the second pre-test. However the inter-individual differences are smaller in the second pre-test, which could be a sign for persistent learning effects. With a second training the same performance as in the first post-test can be achieved.

Second test: differences depending on test stimuli

In the following the results of the second test are presented. Figures 5 and 6 show the results for the pre- and post-tests for both test stimuli. The results of the pre-tests are always plotted on left. With both test stimuli participants show a compressed elevation perception in the pre-tests. The post-tests are plotted on the right and they show an increased ability in the absolute accuracy of elevation perception. When comparing the boxplots of the post-tests, the learning effect seems to be more prominent for the speech-shaped noise. In contrast to the speech signal, increased accuracy is observed for nearly all elevation angles. These results have to be compared carefully, because training time and conditions for both tests were not the same. For the speech stimuli only one session of audio-visual training was conducted, while in advance of the noise test two training sessions were conducted (compare with Fig. 3). On the other hand, all training sessions were conducted with the speech signal and no training with the noise stimuli was carried out.

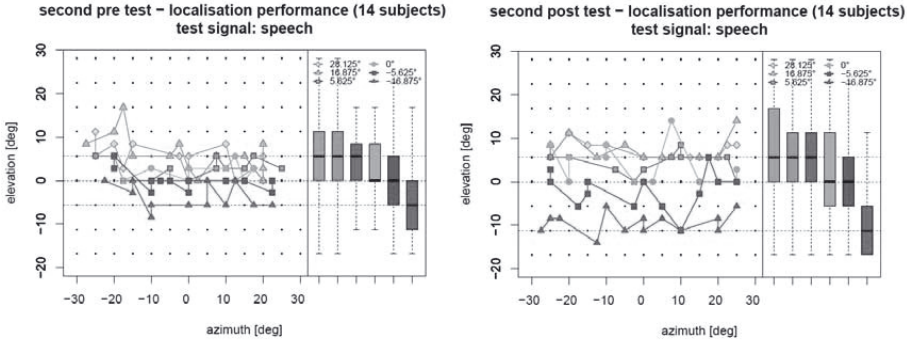

Fig. 5: Median localisation performance for the speech signal in the pre- and post-test; the left side of each plot shows the median ratings for the tested directions and the right side shows the perceived elevation as box plots for each target elevation angle.

All boxplots show a broad range of minimum and maximum values and at some angles high inter-quartile ranges. The reason is found when observing individual results. Depending on the subject the position of the virtual sources has a positive or negative offset (similar to the results of Hofman *et al.* (1998)). After training different learning

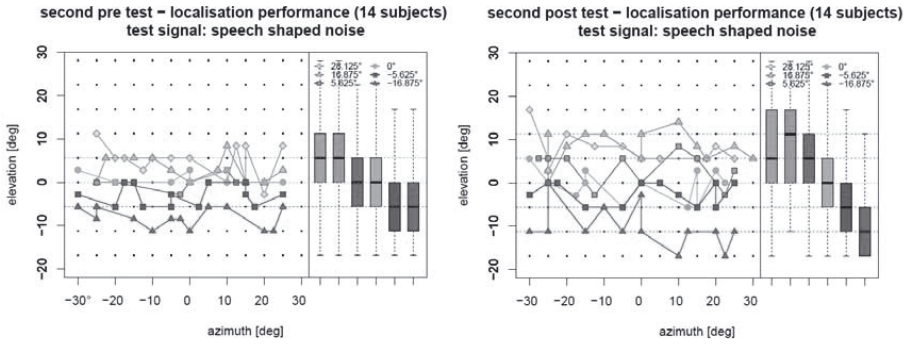

Fig. 6: Median localisation performance for the speech shaped noise signal in the pre- and post-test; the left side of each plot shows the median ratings for the tested directions and the right side shows the perceived elevation as box plots for each target elevation angle.

patterns can also be observed. One participant may increase his performance only in the lower hemisphere and another listener only in the upper hemisphere. Another interesting point are the ratings for the highest elevation angle in post-test for the speech-shaped noise signal (Fig. 6). The highest elevation angle was close to the maximum of the field of view and for some people this row of virtual loudspeaker symbols was barely visible. This might be an explanation for these elevation ratings being out of order.

SUMMARY AND DISCUSSION

As presented in Fig. 4 elevation perception decreases without further training to the initial performance. However smaller inter-individual differences are observable in the second test for people who also participated in the first test. This could mean that, at least for some people, a training effect remains over a longer period of time.

The second test showed that participants were able to increase their localisation performance despite a spectral difference between training signal (speech) and test signal (speech-shaped noise). This could be a sign for perceptual adaptation instead of pattern learning. A clear discrimination between those types of adaptation is still not possible while looking at these results. In the next test iteration, it could be useful to use different training tasks in combination with alternating stimuli. This way it might be possible to exclude pattern learning further. Another advancement would be a comparison to a control group (which gets no training but has to do the test as well) to distinguish between adaptation introduced by the audio-visual training and by the test procedure itself.

Further interesting effects can be observed when looking at the individual results:

Participants show different learning patterns. Performance increases are often observed in only one hemisphere (upper or lower) and the amount of accuracy gain varies highly.

ACKNOWLEDGEMENT

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (Grant BR 1333/14-1).

REFERENCES

Algazi, V.R., Duda, R.O., and Thompson, D.M. (**2001**). "The CIPIC HRTF Database", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 99-102.

Hawkey, D.J.C., Amitay, S., and Moore, D.R. (**2004**). "Early and rapid perceptual learning," Nature Neurosci., **7**, 1055-1056.

Hofman, P.M., Van-Riswick, J.G., and Van Opstal, A.J. (**1998**). "Relearning sound localization with new ears," Nature Neurosci., **1**, 417-421.

Klein, F., and Werner, S. (**2013**). "HRTF adaption under decreased immersive conditions," AIA-DAGA, Meran, Italy, 580-582.

Majdak, P. (**2012**). "Audio-visuelles Training der Schallquellenlokalisation mit manipulierten spektralen Merkmalen," DAGA, Darmstadt.

Parseihian, G., and Katz, B.F.G. (**2012**). "Rapid head-related transfer function adaptation using a virtual auditory environment," J. Acoust. Soc. Am., **131**, 2948-2957.

Schärer, Z. (**2008**). *Kompensation von Frequenzängen im Kontext der Binauraltechnik*. Master thesis, TU Berlin.