Evaluation of a Child-Appropriate Localization Paradigm

Hark Braren, Janina Fels
Institute for Hearing Technology and Acoustics, RWTH Aachen University, 52074 Aachen
Email: {hbr,jfe}@akustik.rwth-aachen.de

Introduction

Virtual acoustic environments presented over headphones are nowadays commonly used in acoustic experiments, including for children [1, 2, 3]. These often utilize non-individual head-related transfer functions (HRTFs), for instance from artificial heads, that are known from experiments with adults to affect, among other factors, the localization accuracy. While objective, including model-based evaluation metrics of child and adult HRTFs show comparable dissimilarities with artificial-head HRTFs [4, 5], subjective evaluations to validate these findings have yet to be done with children.
Previous studies evaluating the localization abilities of children [6, 7] and even infants [8] showed a continuous improvement with age up into their teen years. Hence, when evaluating the effect of non-individual HRTFs on these age groups, a baseline measurement using real world sound sources should be considered to evaluate the localization degradation due to the HRTF and not the overall performance. Another factor that can be considered is the localization of elevated sources. Previous studies in general were limited to lateralization performance in the horizontal plane.
To evaluate effects of non-individual HRTFs in children, an age appropriate 3D-pointing paradigm was designed and implemented amd tested with adults.

Experiment Design

The chosen experimental design can best be described as an identification task, that incorporates gamification features adapted from [9] was chosen. The SCaLAr loudspeaker array [10] (see Figure 1) was chosen for the experiment, as its 68 loudspeaker can be used on the one hand as real world sources and for the in-situ measurement of individual HRTFs. This allows the comparison of real world sources with individual and non-individual HRTFs. The loudspeakers also act as targets in the identification task.
As an easy to understand pointing method, a Logitech presenter was set up with a tracking body for an Optitrack infrared tracking system. The laser pointer in the presenter was modified to be always on. This meant, that the participants simply have to point the laser at a loudspeaker, getting visual feedback from the laser point and press a button on the presenter, as seen in Figure 3. The position and orientation of the presenter is captured using the tracking software, from which a virtual representation of the lasers line can be constructed.
One drawback of the used array is its coarse sampling with approximately 25° between adjacent speakers. Thus the measured HRTF is to coarse for a continuous

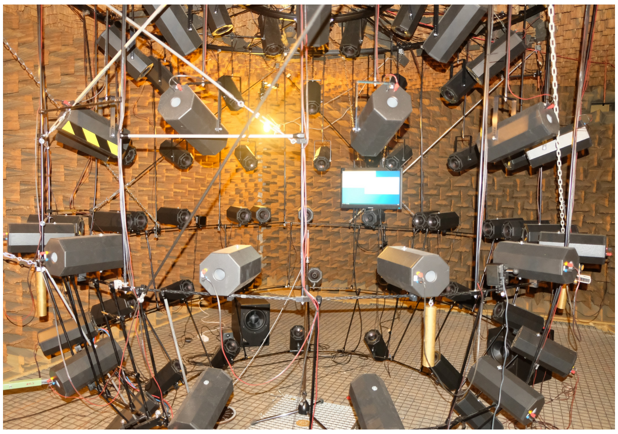

Figure 1: SCaLAr loudspeaker array [10] used for real world targests and in-situ measurements of individual HRTFs

presentation method, where the head of the participants is free to move around, his/her head is tracked and the virtual loudspeaker position/HRTF is continuously updated to have it appear to be in a fixed position. Instead only a static reproduction where the participant has to maintain a set orientation during the presentation of stimuli has to be used.

Before the start of the actual experiment, an audiometry is performed to check for normal hearing. Next, the participants are guided into the array, where their individual HRTF is measured from all loudspeakers in the array. The measurement is carried out using the interleaved swept-sine method [11, 12] and Sennheiser KE3 microphones placed in the ear canal entrance using hearing-aid domes, which block the ear canal. Finally, the transfer functions of the Sennheiser HD650 headphones used for the headphone reproduction is measured using the same microphones in accordance with [13]. Once these measurements are completed, the headphones are removed and the participants got an introduction into the experimental setup and accompanying UI. The instructions are both displayed on the screen in front (c.f. Figure 1) and read aloud by a friendly narrator presented though a loudspeaker directly in front of the participants. During this extensive introduction, the input method, using the Logitech presenter is tested for several source positions using both presentation methods, i.e. headphones and real loudspeakers. Afterwards the actual experiment started, always with the real world loudspeakers as the first block, and a counterbalances set of the three HRTF presentation methods afterwards.

Each block consisted of 16 trials from a given HRTF.

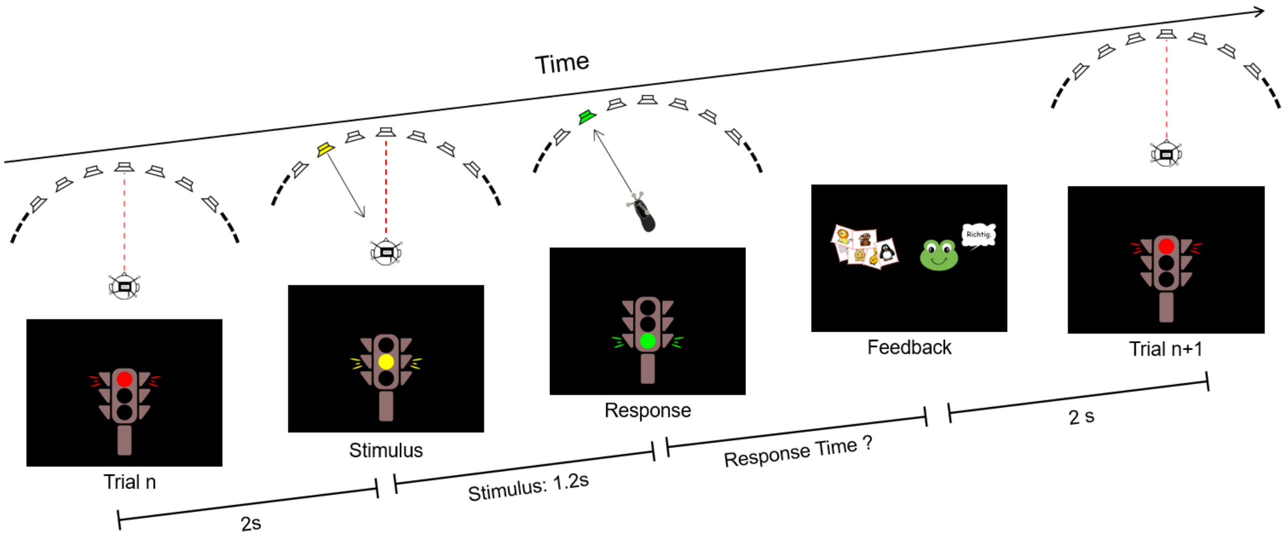

Figure 2: Time diagram of an individual trial in the experiment.

The trials in each block again are counterbalanced using latin squares to avoid order effects. Each trial follows the scheme presented in Figure 2. The participants are presented with a traffic light on the monitor in front. During the red light, they are instructed to look in front at the display and keep a steady and neutral position. A tracking body mounted to the headphones is used to record the position and orientation of the head at this point in time. Next the light turns yellow and the stimulus is presented. The participants are instructed to remain stationary for the duration of the presentation, because of the static presentation method. They are only allowed to move around once the stimulus presentation finished and the light turns green. Then they are instructed to look for the loudspeaker which they heard the stimulus coming from and indicate at it using the tracked presenter in their hand. Their input is confirmed with a audible beep and the participant's attention is directed back at the monitor where they received feedback on their accuracy before continuing with the next trial.

Figure 3: Participant in the SCaLAr array during the experiment.

Validation Experiment

In the presented experiment, 11 adults, age 18-34 (mean 28.3) participated in the experiment as describe above. For the binaural presentation over headphones, the individually measured HRTFs were compared with the KEMAR and ITA artificial head HRTFs. Eight directions in left hemisphere were tested with four repetitions per direction. As the stimulus, a pulsed pink noise consisting of three noise pulses with a overall duration of 1.2 s was chosen and presented at 65 dB across all presentation methods.
The experiment including the audiometry and measurements in the beginning took about 60 min per participant.

Results

First of all, an evaluation of the response time showed no significant differences between the different presentation methods. The average time between the end of the stimulus presentation and the selection of the source by the participant by pressing a button on the presenter as seen in Figure 4. The median response time was 1.6 s for the real sources, 1.8 s for the ITA artificial head, 1.7 s for the KEMAR and a slightly better 1.5 s for the individually measured HRTF.

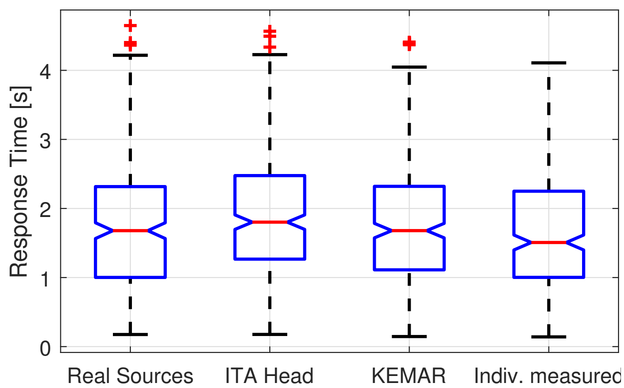

Figure 4: Response time for the different presentation methods.

An evaluation of the head movement between the red traffic light phase and the moment before the light turns green revealed, that the participants moved their head more than anticipated. The translatory motion was on average on average 2.1 cm, with some participants moving their head up to 7 cm. The rotatory motion is presented in Figure 5. Here especially the pitch angle is to be mentioned with up to 7° degrees of motion. This can be attributed to the locations of the tested loudspeakers, with three of the positions being in the saggital plane at different elevations.

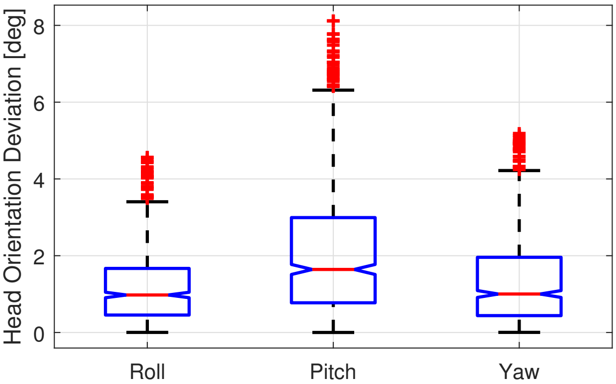

Figure 5: Magnitude of the head rotation angles between the red traffic light and the moment before it turns green.

An evaluation of the 3D-angle between the target and response loudspeaker positions is presented in Figure 6. While the median in the boxplot is the same at 24° for all three HRTFs due to the quantized positions inherent in the identification paradigm, a comparison of the mean values and the distribution of the datapoints indicates, that the individual HRTF performs better compares to the other two in both front-back. The mean angular error values are 9.6° for the real sources, 28° for the ITA artificial head, 28.3°for the KEMAR, and 21.5° for the individually measured HRTF. A comparison of these means using bonferroni correction shows only the difference between the ITA and Kemar artificial head to be not significant. The differences between the real sources and all HRTF types as well as between the individual and non-individual HRTFs are found to be significant. This confirmed previous findings, where HRTFs in general perform worse than the real sources, especially when only using a static reproduction method, and individually measured HRTFs outperforming non-individual ones.

The results from the identification task using the real world loudspeakers once more indicates the limitations of the static reproduction method, i.e. no head-movement is allowed during the stimulus presentation which usually aides localization accuracy. The azimuth error for this task is presented in Figure 7. Even though the spatial sampling of the array is relatively coarse, some loudspeaker locations were harder to localize than others. These were the loudspeaker furthest behind the the participant at 120° azimuth and the most elevated source at 52° elevation and 65° azimuth. Both were located further forward, indicating a potential back-front reversals in the first case.

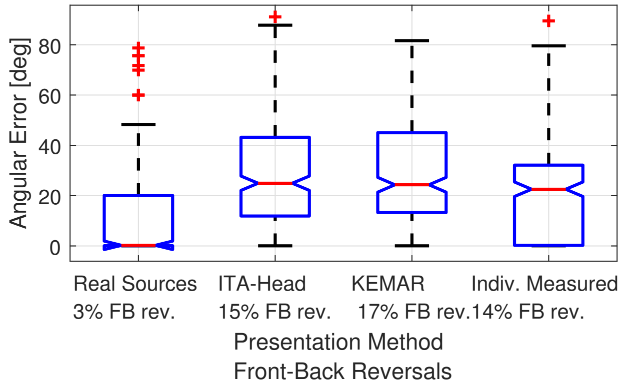

Figure 6: Angular error and percentage of front-back or back-front reversals (FB rev.) for the different presentation methods.

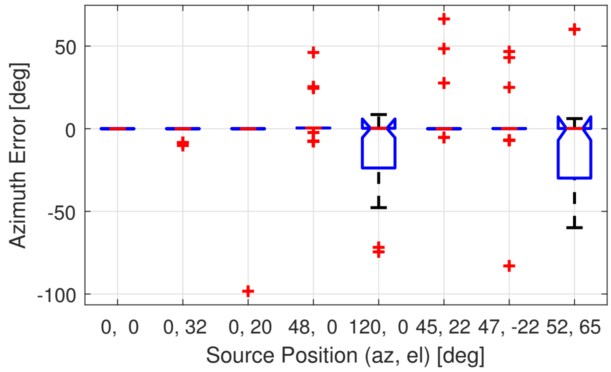

Figure 7: Azimuth Error across all participants using the real loudspeakers as the presentation method (no headphones).

Conclusion

The presented child-appropriate paradigm has the advantage of comparing both real world sources and individually measured HRTFs with non-individual HRTFs from known artificial heads. The downside of the coarse spatial sampling of the SCaLAr array only allows a static reproduction over headphones, which does affect the overall performance. The evaluation of the head movement during the presentation of the stimuli indicates, that participants start moving their head a bit before the stimulus finished, which again can affect the results, as in the static presentation, the virtual sound source moves along with the head. To reduce this effect, the addition of an adjustable headrest can help to keep the head in a more fixed position during the stimulus presentation.
The results of the localization accuracy and front-back reversals are in line with previous studies, and thus what was expected.

References

[1] Janina Fels, Frans Coninx, and Wolfgang H Döring. Application of Binaural Technology in an Adaptive Auditory Speech Test for Children. In *NAG/DAGA*, pages 188–191, Rotterdam, 2009.

[2] Sharon Cameron, Harvey Dillon, and Philip Newall. Development and evaluation of the listening in spa-

tialized noise test. *Ear and Hearing*, 27(1):30–42, 2006.

[3] Daniel Johnston, Hauke Egermann, and Gavin Kearney. Measuring the behavioral response to spatial audio within a multi-modal virtual reality environment in children with autism spectrum disorder. *Applied Sciences (Switzerland)*, 9(15), 2019.

[4] Janina. Fels, Pia Buthmann, and Michael Vorländer. Head-related transfer functions of children. *Acta Acust. united with Acust.*, 90(5):918–927, 2004.

[5] Hark Braren and Janina Fels. Objective differences between individual HRTF datasets of children and adults. In *Proc. Int. Congr. Acoust.*, Aachen, 2019.

[6] S. Kühnle, A. A. Ludwig, S. Meuret, C. Küttner, C. Witte, J. Scholbach, M. Fuchs, and R. Rübsamen. Development of auditory localization accuracy and auditory spatial discrimination in children and adolescents. *Audiol. Neurotol.*, 18(1):48–62, 2012.

[7] Lieselot Van Deun, Astrid Van Wieringen, Tim Van Den Bogaert, Fanny Scherf, F. Erwin Offeciers, Paul H. Van De Heyning, Christian Desloovere, Ingeborg J. Dhooge, Naïma Deggouj, Leo De Raeve, and Jan Wouters. Sound localization, sound lateralization, and binaural masking level differences in young children with normal hearing. *Ear Hear.*, 30(2):178–190, 2009.

[8] BA Morrongiello and PT Rocca. Infants' localization of sounds in the horizontal plane: effects of auditory and visual cues. *Child Dev.*, 58(4):918–927, 1987.

[9] Karin Loh, Edina Fintor, Sophie Nolden, and Janina Fels. Comparing Intentional Switching of Auditory Selective Attention in Children and Adults in an Experiment Suited for Children. *Proc. 23rd Int. Congr. Acoust.*, (1):3677–3678, 2019.

[10] Florian Pausch, Gottfried Behler, and Janina Fels. SCaLAr - A surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments. *Acta Acustica*, 4(5), Oct. 2020.

[11] Piotr Majdak, Peter Balazs, and Bernhard Laback. Multiple exponential sweep method for fast measurement of head related transfer functions. In *Audio Eng. Soc. - 122nd Audio Eng. Soc. Conv. 2007*, volume 2, pages 609–615. Audio Engineering Society, jul 2007.

[12] Pascal Dietrich, Bruno Masiero, and Michael Vorländer. On the optimization of the multiple exponential sweep method. *J. Audio Eng. Soc.*, 61(3):113–124, 2013.

[13] Bruno Masiero and Janina Fels. Perceptually robust headphone equalization for binaural reproduction. In *Audio Eng. Soc. 130th Conv.*, pages 1–7, 2011.