Auditory Training of Spatial Processing in Children with Hearing Loss in Virtual Acoustic Environments: Pretest Results

Z. Ellen Peng, Florian Pausch, Janina Fels

Institute of Technical Acoustics, Medical Acoustics Group,
RWTH Aachen University, Kopernikusstr. 5, 52074 Aachen, Germany, Email: zpe@akustik.rwth-aachen.de

Abstract

Improved efficacy of auditory training was demonstrated in studies by Cameron and Dillon (2007-2012) on normal hearing children with spatial processing disorder, also common among those with hearing loss, by presenting speech training materials in virtual acoustic environments (VAE) with head-related transfer functions (HRTFs) measured from an adult artificial head. To further investigate auditory training using complex realistic acoustic scenes among children fitted with hearing aids, such paradigms are adopted in German for testing in reverberation simulated in loudspeaker-based VAE. In the pretest experiment, while fully immersed in VAE using individualized HRTFs, children are asked to repeat sentences spoken by the target talker located at azimuth 0° when two distractor talkers are simultaneously telling irrelevant stories. Each child undergoes testing in a total of eight conditions, consisting of manipulations in spatial cues (target-distractor collocated vs. spatially separated at 90°), talker pitch cues (target-distractor sharing the same vs. different voice), and room acoustics (0.4 s vs. 1.1 s reverberation time). The speech reception threshold (SRT) is measured adaptively at 50% intelligibility under each condition. The benefits of spatial cues and talker pitch cues are compared between reverberant conditions.

Introduction

Approximately 32 million children worldwide suffer from hearing loss [1]. While assistive hearing devices such as hearing aids provide partial restoration to hearing, spatial processing disorder (SPD) often remains as a co-existing condition with reduced hearing sensitivity [2]. Recently, an auditory training paradigm developed by Cameron and Dillon [3], [4] succeeded in treating SPD among normal-hearing children between 7 and 11 years old. The original paradigm focuses on training the use of spatial cue in a “cocktail party” listening situation, where children with SPD learn how to selectively attend to the target talker and ignore the distractors located in different spatial locations. Pre- and post-tests showed significant improvement in speech intelligibility in anechoic environment.

To provide access for children with hearing loss who are fitted with hearing aids, a project is underway to implement the original paradigms in virtual acoustic environments (VAE) that deliver immersive listening experiences for training under realistic reverberation conditions. This paper discusses the preliminary results from the pretest assessment prior to auditory training.

Methodology

System Description

To provide realistic listening scenarios in VAE, a rectangular room of 270 m³ was simulated in the Room Acoustics for Virtual Environments (RAVEN) software [5] to be outfitted with furniture and surface materials typical for a classroom. Two reverberation time (RT) conditions were created by utilizing different combinations of surface materials, such as carpet, plaster, and acoustic absorber, as shown in Figure 1. The mid-frequency averaged RT was 0.4 s for the low reverberation condition and 1.1 s for the high reverberation condition.

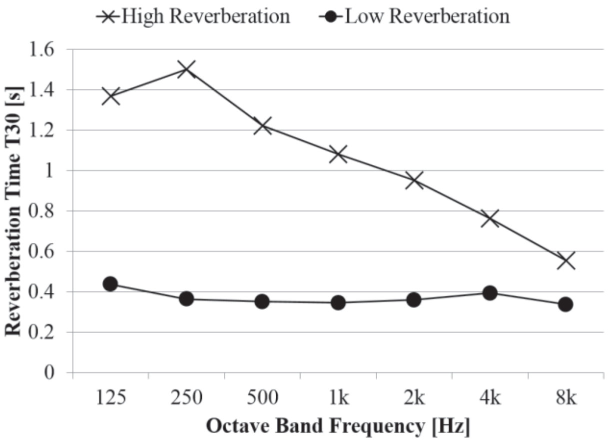

Figure 1: Reverberation time (T30) in seconds as a function of octave band frequency in Hz from the simulated classroom for the low- and high-reverberation conditions.

Individualized head-related transfer functions (HRTFs) and hearing-aid-related transfer functions (HARTFs) were utilized to create binaural room impulse responses, which are updated in real-time accordingly to the child’s head orientation in the virtual scene. It was later determined that individual measurements of HRTFs were deemed impractical with younger children in groups. As a result, an individualization procedure was utilized to modify both interaural-time difference and spectral characteristics [6], [7] in the HRTFs, and only ITD cues in the HARTFs measured from the ITA dummy head [8]. A database of HRTFs and HARTFs was pre-calculated based on typical children’s head dimensions (i.e., height, width, and depth) to cover the 15th, 50th, and 85th percentiles. The sets of HRTFs and HARTFs with nearest dimensions are selected for each child based on the measured head size.

During the experiment, the children will be individually immersed in the VAE created by a 4-channel loudspeaker array with digital cross-talk cancellation filters inside a sound attenuated listening booth [9]. Additional audio playback for children with hearing loss is introduced via a pair of bilateral dummy hearing aids (DHA) with open dome fitting. The DHA receives amplified signals from a master hearing aid software platform [10], which utilizes a built-in fitting procedure to calculate the frequency-dependent insertion gain with dynamic range compression based on individual audiograms using the Cambridge formula [11]. The use of a DHA provides uniform signals to children with hearing loss by eliminating the potential bias from a variety of different signal processing strategies in their individually fitted hearing aids. To ensure safe playback level, additional software and hardware limits were set to avoid peak level exceeding 105 dB SPL (re 20µPa) from the DHA. To allow the use of residual hearing for children with hearing loss, the loudspeaker array is still in use for them to provide a more authentic listening situation. The total system latency, including CTC, MHA, and tracker processing, was restricted to just under 50 ms to ensure a real-time playback environment.

Experiment

In this study, the Cameron and Dillon [3] paradigm of assessing spatial processing disorder (SPD) was adopted to provide assessment in the pre-tests by using new speech materials in German and presenting them in reverberant conditions. In the original paradigm, a total of four conditions are tested, which consist of 2 spatial cues (target-distractor collocated vs. spatially separated at 90°) X 2 talker pitch (target-distractor sharing the same vs. different voices) cues as illustrated in Figure 2.

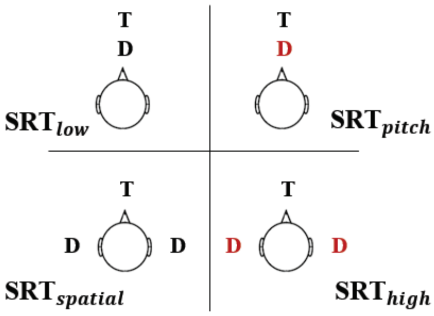

Figure 2: Test conditions of the Cameron and Dillon [3] paradigm for assessing spatial processing disorder. Target talker is always located at azimuth 0° (front of the child). Distractors are either collocated with the target talker (upper row) or spatially separated from the target (lower row) at +/- 90° azimuth from the child. The distractors either have the same voice as the target talker (left column) or different voices from two females other than the target talker (right column). The speech reception threshold (SRT) is measured in each condition.

To emulate a "cocktail party" listening situation, the children were asked to orally repeat the sentences spoken by the target talker, while two distractors are simultaneously telling irrelevant stories. For the speech materials, a subset of the HSM sentences [12] containing only 4- and 5-word sentences was selected as the target stimuli and recorded with a native German-speaking female. Eight unfamiliar Grimm stories in German were recorded with the target talker, as well as two other females who are also native German-speakers. Descriptions of the speaker characteristics are shown in Table 1.

Table 1: Fundamental frequency and speech rate for all speakers in this study

	Fundamental Frequency	Speech Rate
Target	213 Hz	3.4 syllables/s
Distractor 1	191 Hz	3.2 syllables/s
Distractor 2	198 Hz	3.3 syllables/s

All speech materials were recorded anechoically at a 16-bit resolution and 44.1 kHz sampling rate with a Zoom H6 hand-held recording device and a Neumann TLM 170 studio recording microphone. Speakers were instructed to speak with conversational style. All recordings were normalized per EBU-R128 standard [13], which provides normalization based on equal loudness instead of the conventional root-mean-square method of equal energy.

In this study, this paradigm is tested for each child in the 0.4 s and 1.1 s reverberation time conditions to assess children's ability to utilize spatial and pitch cues in realistic room acoustic scenarios.

During the experiment, the distractors are constantly speaking at 55 dB SPL (re 20µPa). A one-up-one-down adaptive procedure is used to measure the speech reception threshold (SRT) of 50% speech intelligibility by changing the target speech level, starting from 70 dB SPL. The step size is set at an initial 4 dB descending until the first reversal, at which point the signal-to-noise ratio (SNR) reduces below the child's SRT, and changed to steps of 2 dB thereafter. Each test block terminates once seven reversals are reached. The SRT is then calculated by averaging the SNRs of the last four reversals. If the child is unable to accurately repeat at least half of the words in the initial sentence, the ascending step size is set at 2 dB. To ensure safe playback level, the test block terminates if the child fails to correctly identify at least half of the words in any of the first five sentences. Moreover, the target level never exceeds 80 dB SPL.

A nested Latin square was utilized to counterbalance the eight test conditions (2 spatial cues X 2 pitch cues X 2 RT). HSM sentences are randomly assigned to match each test condition, where the first 23 sentences in each test block are always 5-word sentences and 4-word sentences thereafter. The distractors' location (left vs. right) and story assignments are also counterbalanced. A 3-second leading of story playback is presented prior to the first target sentence in each test block. A leading beep of 1 kHz pure tone of 200 ms is provided 500 ms prior to each target sentence playback.

Participants

All recruitment was performed after obtaining ethical approval by the Medical Ethics Committee at the RWTH Aachen University (EK 188/15). Twelve children, between 7 and 13 years old, with hearing loss were recruited from the David-Hirsch-Schule in Aachen through teacher-parent communications. These children are all currently fitted with bilateral hearing aids and have severe to profound hearing loss in one or both ears.

Although practice trials were provided, only eight children were able to provide valid data by performing the task in more than six of the eight test blocks. Children who failed to perform the task were removed from data analysis. Among the eight children tested, some were unable to complete 1-2 test blocks, which was also excluded from analysis. In the case of profoundly deaf children, who were unable to receive any stimulation from the MHA, they were asked to utilize their own hearing aid and HRTFs in the loudspeaker-array playback were replaced by HARTFs to correct for additional distance (approximately 2.5 cm) from ear canal entrance to hearing aid microphone position.

Prior to experiment, the children were given a group introduction about the experimental set-up and procedure. Pure tone audiogram for each child was performed by an audiologist within three weeks of the experiment date and was used to create the fitting in the master hearing aid software. Each child received 10 €, deposited directly to their parents' bank account after the experiment.

Results

Simple paired t-test was conducted to compare the SRTs for pitch cue, spatial cue, and reverberation conditions, as plotted in Figures 3-5. There was no statistically significant reduction of SRTs when pitch cue was introduced, $t(28) = 0.45$, $p = .33$. The difference in SRT was found significant for the spatial cue [$t(29) = 3.23$, $p = .002$] and for the reverberation [$t(28) = -3.11$, $p = .002$].

The release from masking (RM) can be calculated using Equation (1) for each RT condition.

$$RM = SRT_{low} - SRT_{cue} \qquad [\text{dB}] \qquad (1)$$

SRT_{cue} is the speech reception threshold in pitch, spatial, or high (both pitch and spatial) cue condition.

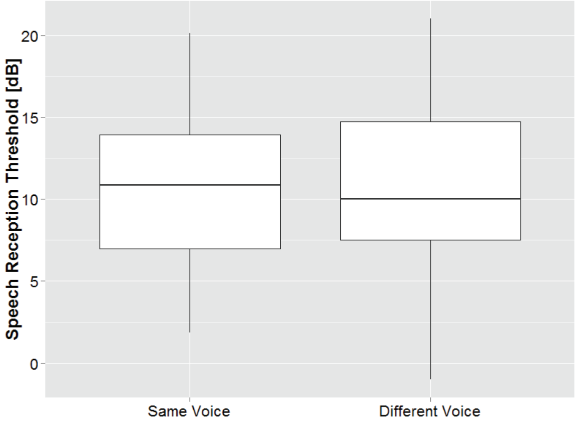

Figure 3: Speech reception threshold in dB in the pitch cue conditions. Distractors either share the same voice or different voice with the target talker. No statistically significant difference was found.

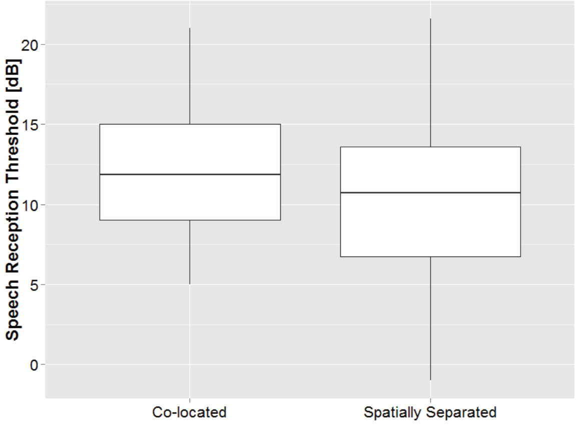

Figure 4: Speech reception threshold in dB in the spatial cue conditions. Distractors either co-located with or spatially separated at ±90° from the target talker.

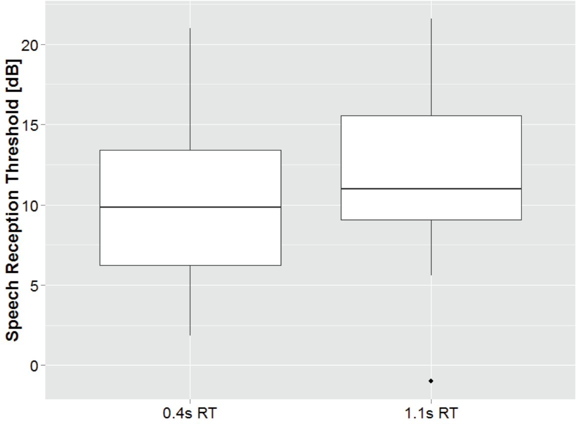

Figure 5: Speech reception threshold in dB in the reverberation conditions of 0.4 s and 1.1 s reverberation time.

A paired t-test is conducted to compare RM for 0.4 s versus 1.1 s RT. Results show a statistically significant difference between RMs in the high cue condition for the low- versus high-reverberation conditions, t(7) = -3.96, p = .004. The RM is greater for 1.1 s than for 0.4 s RT. However, the eight tested children exhibited vast individual differences in response to RT, as shown in Figure 6.

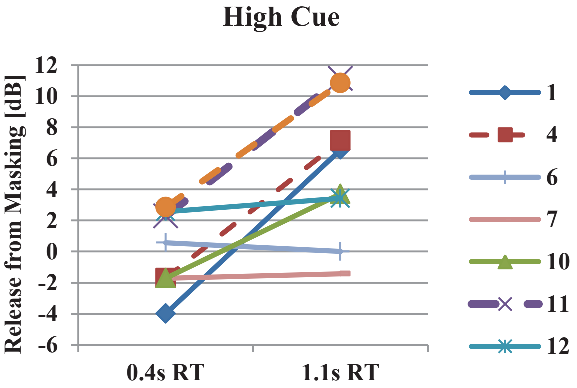

Figure 6: Spatial release from masking for individual children with hearing loss under 0.4 s and 1.1 s reverberation times.

Conclusion and Discussion

In this study, a paradigm to assess spatial processing abilities was adapted for testing children fitted with bilateral hearing loss in virtual acoustic environments. The paradigm tested these children's abilities to utilize spatial cues for improving speech intelligibility, as measured by spatial release from masking (RM), in reverberant environments.

Results suggest that the tested children were able to take advantage of the spatial cue, but not pitch cue, available in a release from masking paradigm implemented in virtual acoustic environments. In comparing the effect of reverberation, RM was larger in the high reverberant condition with 1.1 s than 0.4 s. Large individual differences were observed in responding to the negative effect of reverberation. These individual children's RMs will be used as baseline measures for comparisons after they complete the auditory training in three-month's time.

Acknowledgement

This work received funding support from the European Union's Seventh Framework Programme: Improving Children's Auditory Rehabilitation (iCARE, ITN FP7-607139). The authors would also like to acknowledge Lukas Aspöck and Ramona Bomhardt for providing support in the technical development, as well as student research assistants Arndt Brandl, Karin Loh and Alokeparna Ray for assisting paradigm design and data collection.

Reference

[1] World Health Organization, 2012, URL:
`http://www.who.int/pbd/deafness/estimates`

[2] Glyde, H., et al: The effects of hearing impairment and aging on spatial processing. Ear and hearing 34.1 (2013), 15-28

[3] Cameron, S., and H. Dillon: Development of the listening in spatialized noise-sentences test (LISN-S)." Ear and hearing 28 (2007), 196-211

[4] Cameron, S., H. Glyde, and H. Dillon: Efficacy of the LiSN & Learn auditory training software: randomized blinded controlled study. Audiology research 2 (2012)

[5] Lentz, Tobias, et al.: Virtual reality system with integrated sound field simulation and reproduction. EURASIP journal on applied signal processing 1 (2007), 187-187

[6] Bomhardt, R., and J. Fels.: Analytical Interaural Time Difference Model for the Individualization of Arbitrary Head-Related Impulse Responses. Audio Engineering Society Convention 137. Audio Engineering Society, 2014.

[7] Middlebrooks, J.: Individual differences in external-ear transfer functions reduced by scaling in frequency. The Journal of the Acoustical Society of America 106 (1999), 1480-1492

[8] Schmitz, A. "Ein neues digitales Kunstkopfmeßsystem." *Acta Acustica united with Acustica* 81.4 (1995): 416-420.

[9] Aspöck, L., et al.: Dynamic real-time auralization for experiments on the perception of hearing impaired subjects. Proceedings of AIA-DAGA 2015 Conference on Acoustics, Nuremburg

[10] Grimm, Giso, et al.: The master hearing aid: a PC-based platform for algorithm development and evaluation. Acta acustica united with Acustica 92 (2006), 618-628

[11] Moore, B., and B. Glasberg.: Use of a loudness model for hearing-aid fitting. I. Linear hearing aids. British journal of audiology 32 (1998), 317-335

[12] Hochmair-Desoyer, I., et al.: The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users. The American journal of otology 18 Suppl (1997), S83-S83

[13] EBU-R128: Loudness normalisation and permitted maximum level of audio signals. Geneva, 2014.