Melissa Ramírez, Johannes M. Arend, Petra von Gablenz, Katharina Schmidt, Karsten Plotz, Christoph Pörschmann

Towards the virtualization of a sound source localization acuity test to aid the diagnosis of spatial processing disorder in school-aged children: An experimental approach

Document type
Conference paper | Published version
(i. e. publisher-created published version, that has been (peer-) reviewed and copyedited; also known as: Version of Record (VOR), Final Published Version)

This version is available at
https://doi.org/10.14279/depositonce-15554

Citation details
Ramírez, M.; Arend, J. M.; von Gablenz, P.; Schmidt, K.; Plotz, K. & Pörschmann, C. (2022): Towards the virtualization of a sound source localization acuity test to aid the diagnosis of spatial processing disorder in school-aged children: An experimental approach. In: AIA-DAGA 2022 : proceedings of the International Conference on Acoustics. Berlin: Deutsche Gesellschaft für Akustik e.V. pp. 282-285.

Towards the virtualization of a sound source localization acuity test to aid the diagnosis of spatial processing disorder in school-aged children: An experimental approach

Melissa Ramírez[1,2], Johannes M. Arend[2,1], Petra von Gablenz[3], Katharina Schmidt[3], Karsten Plotz[3],
Christoph Pörschmann[1.]

[1] *Institute of Communications Engineering, TH Köln – University of Applied Sciences, D-50679 Cologne, Germany*
[2] *Audio Communication Group, Technical University of Berlin, D-10587 Berlin, Germany*
[3] *Inst. of Hearing Technology and Audiology, Jade University of Applied Sciences, D-26121, Oldenburg, Germany*
E-Mail: melissa.ramirez@th-koeln.de

Introduction

Spatial hearing is an essential auditory function. It allows us to localize, segregate, and group sound sources in space [1], [2]. Accurate sound source localization is a fundamental ability for understanding and following speech in everyday situations, as it contributes to our capacity to discern between target signal streams and other simultaneous sound sources that can be regarded as noise (cocktail party processing) [3].

The spatial processing disorder (SPD) is an auditory processing disorder (APD) characterized as a reduced ability to use binaural cues to achieve spatial release from masking (SRM) despite (but not necessarily) having normal hearing thresholds [4], [5]. When binaural cues are available, the listener can have an SRM in the range of 13-15 dB [6]. However, a person with SPD cannot effectively utilize those binaural cues, requiring a much higher signal-to-noise ratio (SNR) to achieve the same speech reception thresholds (SRTs) than someone with typical spatial processing abilities.

Previous research has found a relationship between the presence of SPD and history of conductive hearing loss due to otitis media in early childhood. This fluctuating (and) or asymmetrical hearing loss, occurring when the brain is developing the skills required to process binaural cues, may disrupt the brain's ability to combine and interpret the spatial information encoded in the interaural time and -level differences (ITDs and ILDs) between the left and right auditory pathways, and children may develop SPD as a result [7]. Children with SPD require thus approximately 4 dB more of signal in a classroom-like listening situation to achieve the same SRM as their peers [6]. They have to allocate a disproportionate amount of resources to the basic signal encoding tasks, which results in a decreased availability of resources for higher-order functions such as comprehension and memory [8]. Therefore, they often have problems concentrating and learning or understanding language and, consequently, reading and writing [9].

International guidelines for the diagnosis of auditory processing disorders and recent literature have already extensively documented the need to include binaural localization abilities as part of the comprehensive battery of tests to aid the diagnosis of patients with suspected auditory processing disorders such as SPD. These guidelines recommend measuring binaural localization abilities to assess directional hearing, the development of binaural processing, and evaluating the auditory localization performance of patients with listening difficulties and after hearing aid or cochlear implant fitting, among others [10].

Moreover, even though the evaluation of auditory spatial localization abilities is of great importance for different patient age groups, there has been an upsurge in clinical and public awareness of the necessity to develop children-appropriate procedures that can be widely used in clinical practice; particularly since recent research has found extensive evidence supporting that remediation can be offered as auditory training, which has shown to be effective for children with auditory processing disorders [11], [12]. Hence, offering early and widely available diagnoses is vital to facilitate access to deficit-specific, appropriate, and effective remediation for children with listening difficulties on time.

The ERKI method [13] (ERKI – from German "Erfassung des Richtungshörens bei Kindern," which translates as "acquisition of directional hearing in children") was developed to close this gap and as an effort to embed directional hearing measurements into typical clinical audiological apparatus and procedures. It measures the listener's sound source localization abilities over the horizontal plane with an approved medical device of the same name, offering a higher spatial resolution and improving the procedure's ease of use and automatability if compared with similar devices previously available. However, even though suitable technical solutions are available that operate with basic hardware equipment, binaural hearing abilities are still not assessed widely as a typical audiological practice.

Digital healthcare is rapidly evolving, and virtual- and augmented reality (VR and AR) technologies have made medical training, diagnosis, and treatment more portable, widely accessible, and less expensive. Hearing healthcare is not the exception [14], and implementing the ERKI method in VR could open new possibilities. It could make the system more portable and accessible, make the measurement procedure more engaging for children, and allow to exploit multisensory interactions between audition, vision, touch, and proprioception that are highly relevant for sound source localization (see [15] for a review). Thus, a VR implementation of the ERKI method could be a tool to assess directional hearing, aiding the diagnosis of APDs such as SPD. Likewise, it could be a tool to investigate the potential of VR on spatial hearing rehabilitation by allowing to include state-of-the-art auditory spatialization techniques and target spatial cross-modal reorganization through multisensory stimulation.

Nonetheless, a challenge arises when pursuing the virtualization of the measurement procedure. The VR application should be able to measure the listeners' binaural sound source localization abilities with comparable performance to the currently available method to prove to be a feasible alternative. However, in the case performance shows to be inferior, it would be a challenge to systematically investigate which test components’ virtualization has a more significant effect on the system’s performance decrease, e.g., if it is due to the new pointing method, to the new headphone-based audio presentation, the use of virtual visual feedback or the use of the head-mounted display. We, therefore, propose a controlled virtualization to address this concern and get a better understanding of the effect that each test component’s virtualization may have on the overall performance of such a VR implementation.

This work presents the conceptualization and development of an experimental setup designed to assess the feasibility of a child-appropriate sound source localization test in virtual reality based on the ERKI method. The setup consists of a series of AR and VR scenarios designed to perform a controlled virtualization of the measurement procedure. Each scenario gives one step toward full virtualization by replacing one test component with its virtual counterpart at a time: First, the pointing method, then the audio reproduction, and finally, the visual presentation until achieving complete virtualization. Thus, the study participants' binaural hearing abilities will be measured in all three proposed scenarios, presented in a randomized order. This way, we can evaluate each scenario's performance separately, making it possible to understand better which test component requires special attention when virtualization is pursued.

Methods

The ERKI method [13] determines the angular localization error (AE) in non-speech sound source localization tasks over the horizontal plane. The setup consists of five loudspeakers arranged in a front semicircle around the subject (0°, ±45°, ±90°; r = 1 m), hidden beneath an acoustically transparent curtain so that neither the amount nor the loudspeakers' position is visible. A LED bar along the semicircle allows the listener to indicate the direction from which they perceive the stimuli (Figure 1). The system can display 37 sound sources (5 real and 32 virtual sources), corresponding to an angular resolution of 5 degrees. The 32 virtual sound sources are generated as phantom sources between adjacent loudspeakers using the vector base amplitude panning method (VBAP) [16].

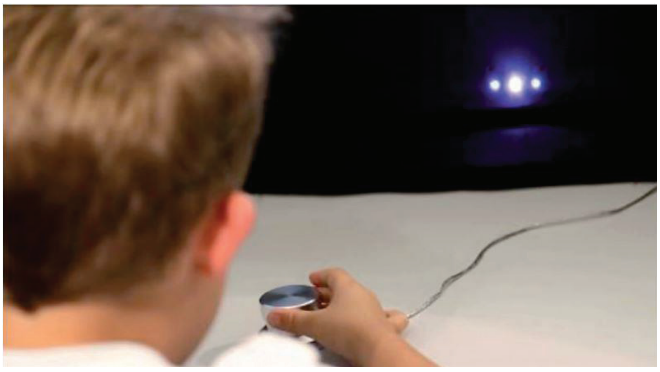

Figure 1: ERKI method (picture taken from [13]). The child sits in the center of the loudspeaker semicircle with the head aligned to the 0° position. A short noise signal (300 ms) at 65 dB SPL is presented. Their task is to determine the direction of the perceived signal using a rotatable switch to position a light at the LED bar. The child presses the button to confirm their selection. Then, a light signal fades in to announce the next stimulus presentation and reminds the participant to re-align the head to the 0° position. All 37 angles from -90° to +90° are randomly presented.

We recreated the ERKI setup in the sound-insulated anechoic chamber of the acoustics laboratory at the TH Köln, which has dimensions of 4.5 x 11.7 x 2.3 m (W×D×H) and a lower cut-off frequency of around 200 Hz. We used five Genelec 8020D loudspeakers as the real sources and generated the remaining 32 virtual sound sources using the VBAP method [16]. We used two consecutive strands of Adafruit WS2801 pixels to display the location of the 37 sound sources as individual LED lights, using a serial peripheral interface (SPI) to transmit the color data and clocked by a microcontroller board (Arduino Mega 2560). An opaque, acoustically transparent fabric was used to cover the setup (Figure 2).

An OptiTrack system with an update rate of 120 Hz was used to track both the listener’s head orientation and the handheld controller used for pointing, ensuring that the stimuli are not reproduced unless the participant's head is aligned to 0° azimuth (central loudspeaker). A 300 ms-long broadband white noise was used as stimuli to avoid the listeners' turning their heads towards the perceived direction during the stimulus delivery (as typical reaction times are around 400 ms [17]).

The test procedure was implemented, controlled, and executed by a purpose-built MATLAB application running on a PC. Moreover, child-appropriate speech instructions were included to improve the setup’s ease of use and gamified character, creating a completely automated measurement procedure.

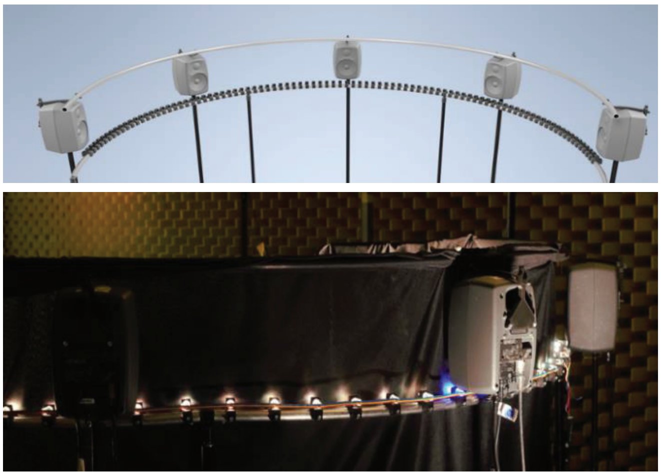

Figure 2: Experimental setup in the anechoic chamber at TH Köln. Top: 3D rendering of the setup without the fabric cover (designed by K. Altwicker, TH Köln). Bottom: Rear view of the setup, including the LED light strip and the acoustically transparent opaque covering.

AR scenario #1: Virtualization of the pointing method

The first test scenario is an augmented reality (AR) scenario aimed to analyze the effect of virtualizing the pointing method as an isolated variable. We used a modified Oculus Quest 2 controller to replace the rotatable switch on the ERKI procedure, changing the pointing method's spatial coding from allocentric to egocentric. With the rotatory switch, the listeners were required to indicate the location of the perceived sound source by spinning the control (which was attached to the table) until locating the LED light in front of the perceived sound source position (Figure 1).

In contrast, with the proposed pointing method, which is also the most used pointing method in VR, the listeners simply extend their arm and point in the direction where they think the stimuli came from (representing the location of the sound object in space relative to their self-body axes). Visual feedback is provided as the LED's color changes as the participants point at them. They press a button in the controller to confirm their answer (Figure 3).

AR scenario #2: Virtualization of the audio reproduction

The loudspeaker-based audio reproduction was replaced with a binaural (headphone-based) presentation in this test condition, which was otherwise identical in design and procedure to the first test scenario (AR scenario #1: Virtualization of the pointing method).

The test subject sits in the center of the semicircle, listens to the stimuli via headphones, and indicates the perceived sound source location in the same way as they did in the previous scenario. We employed measured far-field head-related transfer functions (HRTFs) from a Neumann KU100 dummy head [18] for the binaural presentation. The HRTF set (with a Lebedev 2702 spatial sampling configuration) was transformed to the spherical harmonics (SH) domain at a sufficiently high spatial order of N = 44, allowing artifact-free SH interpolation to obtain HRTFs for the desired direction using the SUpDEq toolbox [19]. The resulting dataset is utilized in the present case to represent the 37 virtual sound sources accurately.

A generic headphone compensation filter was applied to the precomputed stimuli (noise test signal convolved with the respective HRTF) to minimize the influence of the employed headphones (Sennheiser HD 600). The filter is based on 12 measurements (putting the headphones on and off the Neumann KU100 dummy head) to account for re-positioning variability and was designed by regularized inversion of the complex mean of the headphone transfer functions [20] using the implementation by Erbes et al. [21].

Figure 3*:* A test subject during the pilot studies. The listener extends their arm pointing to the location of the perceived sound object in the space relative to their self-body axes (egocentric pointing method).

VR scenario: A virtual version of the ERKI method

The third and last scenario is a fully immersive VR environment. In this case, the LED array was replaced by its virtual counterpart (Figure 4).

Well done! Now you will hear a sound.Where does it come from?

Figure 4: Screen capture of the VR scenario. A child-appropriate, gamified, and completely automated directional hearing measurement application based on the ERKI method was developed to run on a standalone head-mounted display (Oculus Quest 2).

The test subject simply wears the head-mounted display (HMD), the headphones and starts playing the game. The virtual LED array automatically adjusts to the listener's interaural axis height at the beginning of the game. The stimuli are played only when the listener's head is aligned to 0° azimuth and if their head or the HMD is not tilted (roll control). If the head orientation or position is not ideal, voice instructions ask the listener to correct this. Moreover, this scenario integrates dynamic binaural audio spatialization. We evaluated different renderers for the dynamic binaural presentation, and the application is available using the STEAM® Audio SDK [22] and the Unity Wrapper for 3DTI [23], [24]. We employed the same HRTF dataset as in the previous scenario (AR scenario #2: Virtualization of the audio reproduction), but in this case, in SOFA format.

Evaluation and future work

A preliminary feasibility study with normal-hearing young adults is currently ongoing. Its results are used to determine if further improvements need to be made in the study's design and implementation, for effect size calculation and power analysis, among others. Each study participants' binaural hearing abilities are assessed in all three proposed scenarios, which are presented in a randomized order.

The experimental approach proposed in this work allows the evaluation of the feasibility of the VR implementation developed in this study and a systematic evaluation of the effect that each component's virtualization has on the measurement performance. The study results may give insights into which test component requires further improvement, e.g., if other pointing methods shall be implemented or if HRTF individualization approaches are needed. Table 1 summarizes the parameter settings in each test scenario on the proposed experimental approach and compares them with the currently available ERKI method [13].

	ERKI method	Controlled virtualization approach		
		AR #1	AR #2	VR
Sound sources	5 real 32 virtual		37 virtual	37 virtual
Visual feedback	Yes. LED bar	Yes. LED array		Yes. VR environment
Pointing method	Allocentric. Rotatory disc	Egocentric. Modified Oculus Quest 2 controller		Egocentric. Oculus Quest 2 controller
Head tracking	No	Yes (for head orientation control)		Yes (for binaural auralization)
Stimuli presentation	Loudspeaker-based		Headphone-based	
Spatialization technique	VBAP	VBAP	Binaural spatialization (non-individual HRTFs)	Dynamic binaural rendering (non-individual HRTFs)

Table 1. Parameter settings in the ERKI method and all the scenarios developed for the proposed controlled virtualization.

Following studies are planned to evaluate the performance of the VR application developed in this work as a directional hearing measurement tool and the reproducibility of its results.

Conclusion

The present work describes an experimental setup designed to assess the feasibility of a child-appropriate sound source localization test in virtual reality based on the ERKI method. The results of the planned following study may answer whether the performance of the VR application developed in this work to measure the listeners' binaural sound source localization abilities is equivalent or comparable to the method proposed by Plotz and Schmidt [13].

Acknowledgments

We want to give special thanks to Kai Altwicker for his support in designing and printing all the parts required for the measurement setup. Thanks to Miguel Ángel Olivares for developing the VR scenario. Thanks to Raphaël Gillioz for assisting with the visual documentation of the setup.

This work was funded by the German Federal Ministry of Education and Research (13FH666IA6-VIWER-S).

Literature

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Revised Ed. London, UK.: MIT Press, 1997.

[2] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, 1990.

[3] A. W. Bronkhorst, "The cocktail-party problem revisited: early processing and selection of multi-talker speech," *Attention, Perception, Psychophys.*, vol. 77, no. 5, pp. 1465–1487, 2015, doi: 10.3758/s13414-015-0882-9.

[4] S. Cameron and H. Dillon, "Development and evaluation of the LiSN & learn auditory training software for deficit-specific remediation of binaural processing deficits in children: Preliminary findings," *J. Am. Acad. Audiol.*, vol. 22, no. 10, pp. 678–696, 2011, doi: 10.3766/jaaa.22.10.6.

[5] H. Cameron, S., & Dillon, "Remediation of spatial processing issues in central auditory processing disorder," in *Handbook of central auditory processing disorder: comprehensive intervention Vol. 2*, 2nd Ed., F. M. Chermak Gail, Ed. San Diego, CA: Plural Publishing, 2014, pp. 201–224.

[6] H. Glyde, S. Cameron, H. Dillon, L. Hickson, and M. Seeto, "The effects of hearing impairment and aging on spatial processing," *Ear Hear.*, vol. 34, no. 1, pp. 15–28, 2013, doi: 10.1097/AUD.0b013e3182617f94.

[7] K. Graydon, G. Rance, R. Dowell, and B. Van Dun, "Consequences of Early Conductive Hearing Loss on Long-Term Binaural Processing," *Ear Hear.*, vol. 38, no. 5, pp. 621–627, 2017, doi: 10.1097/AUD.0000000000000431.

[8] T. J. Bellis and J. D. Bellis, "Central auditory processing disorders in children and adults," in *The human auditory system: Fundamental organization and clinical disorders.*, Edinburgh: Elsevier, 2015, pp. 537–556.

[9] C. C. Crandell and J. J. Smaldino, "Classroom Acoustics for Children With Normal Hearing and With Hearing Impairment," *Lang. Speech. Hear. Serv. Sch.*, vol. 31, no. October, pp. 362–370, 2000.

[10] F. E. Musiek and G. D. Chermak, *Handbook of Central Auditory Processing Disorder, Volume I: Auditory Neuroscience and Diagnosis*, Second Edi. San Diego, CA: Plural Publishing, 2014.

[11] J. H. Y. Loo, S. Rosen, and D.-E. Bamiou, "Auditory training effects on the listening skills of children with Auditory Processing Disorder," *Ear Hear.*, vol. 37, no. 1, pp. 38–47, 2016, doi: 10.1097/AUD.0000000000000225.

[12] M. Sharma, S. C. Purdy, and A. S. Kelly, "A randomized control trial of interventions in school-aged children with auditory processing disorders," *Int. J. Audiol.*, vol. 51, no. 7, pp. 506–518, 2012, doi: 10.3109/14992027.2012.670272.

[13] K. Plotz and K. K. Schmidt, "Lokalisation realer und virtueller Schallquellen mit einem automatisierten Erweiterungsmodul am Mainzer-Kindertisch – Entwicklung des ERKI-Verfahrens," *Zeitschrift für Audiol.*, vol. 56, no. 1, pp. 6–18, 2017.

[14] J. Murphy, "Virtual Reality: The Next Frontier of Audiology," *Hearing Journal*, vol. 70, no. 9, pp. 24–27, 2017.

[15] A. Maravita, C. Spence, and J. Driver, "Multisensory integration and the body schema: Close to hand and within reach," *Curr. Biol.*, vol. 13, pp. 531–539, 2003, doi: 10.1016/S0960-9822(03)00449-4.

[16] V. Pulkki, "Virtual Sound Source Positioning Using VBAP," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.

[17] G. J. P. Savelsbergh, J. B. Netelenbos, and H. T. A. Whiting, "Auditory Perception and the Control of Spatially Coordinated Action of Deaf and Hearing Children," *J. Child Psychol. Psychiatry*, vol. 32, no. 3, pp. 489–500, 1991, doi: 10.1111/j.1469-7610.1991.tb00326.x.

[18] B. Bernschütz, "A Spherical Far Field HRIR HRTF Compilation of the Neumann KU 100," in *Proceedings of the 39th DAGA*, 2013, pp. 592–595, [Online]. Available: https://zenodo.org/record/3928297#.YkRyyChBxhE.

[19] C. Porschmann, J. M. Arend, and F. Brinkmann, "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 6, pp. 1060–1071, 2019, doi: 10.1109/TASLP.2019.2908057.

[20] A. Lindau and F. Brinkmann, "Perceptual evaluation of head-phone compensation in binaural synthesis based on non-individual recordings," *J. Audio Eng. Soc.*, vol. 60, no. 1/2, 2012.

[21] V. Erbes, H. Wierstorf, M. Geier, and S. Spors, "Free Database of Low-Frequency Corrected Head-Related Transfer Functions and Headphone Compensation Filters," in *Proceedings of 142nd Audio Engineering Society Convention, e-Brief 325*, 2017, pp. 1–5.

[22] "STEAM Audio SDK." https://valvesoftware.github.io/steam-audio/.

[23] M. Cuevas-Rodríguez *et al.*, "3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation," *PLoS One*, vol. 14, no. 3, Mar. 2019, doi: 10.1371/JOURNAL.PONE.0211899.

[24] A. Reyes-Lecuona and L. Picinali, "Unity Wrapper for 3DTI." https://github.com/3DTune-In/3dti_AudioToolkit_UnityWrapper.