Binaural hearing in the real and virtual world to improve school-aged children's listening experience (ViWer-S)

Melissa Ramírez[1,2], Ulrik Kowalk[3], Johannes M. Arend[2], Jörg Bitzer[3],

Christoph Pörschmann[1], Karsten Plotz[3]

[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] Institute of Hearing Technology and Audiology, Jade University of Applied Sciences, D-26121, Oldenburg, Germany

keywords: spatial hearing, auditory processing disorder, spatial processing disorder, auditory scene analysis, hearing assistive technology

1. Introduction

Our brain has the ability to combine and interpret the spatial information encoded in both binaural and monaural cues, allowing us to localize, segregate, and group sound sources in space [1, 2]. The ability to accurately localize sound sources is essential for understanding and following speech in everyday situations [3]. It contributes to our capacity to discern between target signal streams and other simultaneous sound sources that are regarded as noise.
Patients with spatial processing disorder (SPD) have 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, listeners are expected to have a spatial listening advantage in noise of 13-15 dB [6]. However, a person with SPD requires a much higher signal-to-noise ratio (SNR) to achieve the same speech reception thresholds (SRTs) than someone with typical spatial processing abilities.
The diagnosis and remediation of auditory processing disorders (APDs) such as SPD are of great relevance for all demographics. However, there has been an upsurge in clinical and public awareness of the need to develop child-appropriate and accessible procedures and technologies to aid both the diagnosis and treatment of these deficits. An important reason for this is their high prevalence in the school-aged population. It is estimated that APDs affect around 5% of school-aged children and half of all children with learning disorders [7–9], but the true prevalence may be greater due to undiagnosed or misdiagnosed cases. Moreover, APDs may hamper academic progress if untreated. In a classroom-like listening environment, children with SPD must allocate a disproportionate amount of resources to the basic signal encoding tasks, resulting in a decreased availability of resources for higher-order functions such as comprehension and memory [10–12].
The wireless acoustic transmission system or so-called "FM system" is the most common hearing assistive device in the classroom. Nevertheless, although SNR-enhancing technology is a necessary first step toward creating viable classroom learning environments, it is insufficient to address the challenges experienced by children with SPD. When using personal FM receivers, as with every monophonic signal, the listener perceives the teacher's voice internalized, i.e., inside or in the middle of their head, for several hours a day. This effect is very unnatural, and the prolonged exposure to non-spatialized auditory stimuli does not promote the improvement of their binaural hearing abilities [9].
To address this concern, the research project "ViWer-S" focuses on developing alternative hearing assistive technologies (HATs) for school-aged children with SPD. The first approach, "Binaural hearing in the classroom," focuses on developing an alternative to the traditional wireless acoustic transmission system. We propose a technical solution that allows real-time dynamic binaural presentation of the classroom's sound field. Additionally, "Binaural hearing in the virtual world," the second approach of our research, focuses on developing applications in virtual reality (VR) to aid the diagnosis of SPD and promote the improvement of the listener's binaural listening abilities through auditory training. This manuscript summarizes the current state and the following work within the ViWer-S research project.

2. Binaural hearing in the classroom

Our first approach to this concern focuses on developing a technical solution that allows real-time dynamic binaural presentation of the classroom's sound field. It consists of a small and portable microphone array that can be placed on the child's table. Driven by a small, portable computer, the software determines the direction of the sound sources in the room, improves the quality of the signals of interest, and transmits them to the child's hearing aid using binaural rendering. Moreover, thanks to the integrated head-tracking device, the presented sound field is updated in real-time according to the listener's head movements.

With the proposed system, we assist SPD-affected children in the classroom by presenting them with a sound field in which the natural process of auditory attention has fewer obstacles to overcome. This is achieved in three steps. First, we identify different sound sources and estimate their direction of arrival (DOA) relative to the child's head orientation. The most relevant sources are extracted through acoustic beamforming (the term "relevant" will be discussed in Sec. 2.1). Next, these monophonic acoustic signals are enhanced so that the amount of background noise is reduced and – in the case of speech – their intelligibility is increased. In the last step, the signals are provided with spatial information, so they are subjectively localized according to their actual position in the room when presented through hearing aids or headphones.

2.1 Methods

DOA estimation and source extraction: We utilize a 15-channel 2-dimensional microphone array with approximate dimensions of 0.4 m x 0.13 m (WxL). The omnidirectional MEMS microphones are double-logarithmically positioned along an arc (Figure 1). We compile acoustic features from signal frames of length 32 ms that are input to a deep neural network (DNN) that estimates the DOA of all active acoustic sources, producing a new set of DOA estimates for every frame. As input features, we use the well-known generalized cross-correlation with phase transform (GCC-PHAT) [13], with the time difference limited to 24 samples. These input features are processed by a DNN, consisting of 3 convolutional layers and 3 fully connected layers, the last being the output layer with 72 neurons. The output neurons represent 72 classes that span the azimuthal range between 0 and 355 degrees, with a 5-degree resolution. The larger the value at an output neuron, the higher the likelihood that sound from an acoustic source is arriving from the respective direction. The maxima resulting from the output layer are treated as DOA estimates, which are then conveyed to a source tracking mechanism. More information on the design and training of the DNN can be found in [14]. Additionally, a super-directive beamformer is used to extract the signal from the sound field using all 15 microphones, resulting in a number of discrete monophonic signals – one for every detected source. A denoising algorithm is applied to every extracted source signal to enhance speech intelligibility.

Apart from the microphone array, we assume the existence of a body-worn support microphone that is close to the teacher's mouth (hereafter called external microphone). We use the external microphone as the source for the teacher's speech, and its signal is exploited to aid the estimation of the teacher's position.

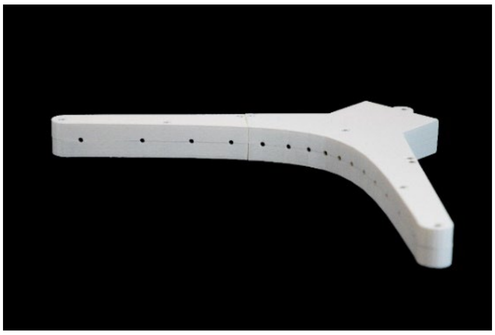

Figure 1. Prototype of the 15-channel 2-dimensional microphone array placed in front of the child, with approximate dimensions of 0.4 m x 0.13 m (WxL).

Signal presentation: A spatial sound field is generated by convolving the discrete, monophonic source signals with head-related transfer functions (HRTFs). After convolution, the new signal is perceived as coming from a specific direction outside the listener's head when presented through headphones or hearing aids. Apart from the extracted sound sources, no other components are considered for the presentation, leaving a sound field that is less perceptually demanding for the listener. We use the information from a head-tracking device to account for the listener's head movements, providing a stable image of the sound sources in the room, even when the head is rotated. All operations must be performed in real-time to provide a realistic listening environment without creating a perceptible delay in either localization or acoustic processing. Therefore, the DNN is designed to be very efficient, and all spatialization and enhancement operations utilize segmented block processing.

Source relevance: The teacher is the most important acoustic source in the classroom. When the teacher is speaking while other sources are active, the teacher is assigned the highest priority in the source management mechanism. Besides, it is imperative to retain which tracked sources are allocated to the teacher as they are most likely to change their position over time. This knowledge is mainly gathered by exploiting the information from the external microphone, primarily dominated by the teacher's speech.

2.2 Preliminary results

Preliminary experiments concerning the performance of the DOA estimation algorithm have shown that the proposed DNN can provide accurate DOA estimates under realistic acoustic conditions [14]. In simulated rooms with classroom dimensions and reverberation, the DNN has produced estimates with a precision of approximately 1.5 degrees for single speakers. Studies regarding the enhancement and spatialization performance, as well as multi-speaker estimation accuracy in real-time, are planned for future work.

3. Binaural hearing in the virtual world

As a second approach, we work on developing a series of virtual reality (VR) applications to make the diagnosis and remediation of SPD more portable, widely accessible, and less expensive. On the one hand, diagnosis is supported by developing applications to assess the binaural listening abilities of children in an automated and gamified way. On the other hand, we investigate the potential of VR on spatial hearing rehabilitation, developing serious games to train spatial hearing abilities, including sound source localization and speech intelligibility tasks.

This way, children with SPD can train and improve their spatial hearing abilities by playing age-appropriate games on a standalone head-mounted display at times of their choice, at home, and under adult supervision. VR technologies allow us to use state-of-the-art auditory spatialization techniques and target spatial cross-modal reorganization through multisensory stimulation, exploiting multisensory interactions between audition, vision, touch, and proprioception that are highly relevant for sound source localization [15].

3.1 Methods

We have recently published the conceptualization and development of an experimental setup designed to assess the feasibility of a child-appropriate sound source localization test in VR based on the ERKI method [16]. The setup consists of a series of augmented reality (AR) and VR scenarios designed to perform controlled virtualization of the measurement procedure [17].

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. With this experimental setup, we can assess the listener's sound source localization acuity with an intuitive egocentric pointing method in three different audio presentation modalities (loudspeaker-based, binaural headphone-based, or inside a VR environment presented over a standalone head-mounted display). We can therefore evaluate and understand each test component's influence on the virtualization process and the feasibility of using VR as a tool to assess spatial hearing abilities. The reader may refer to [17] for technical details and a description of the proposed experimental setup.

3.2 Preliminary results

We conducted a listening experiment with 20 normal-hearing adults to evaluate the performance of each of the AR and VR scenarios proposed in [17] as directional hearing measurement tools in a repeated measure or within-subject study design. A publication with a detailed analysis of the results will follow. It is of particular interest to answer whether the performance of the VR application developed in the previous work to measure the listeners' binaural sound source localization abilities is equivalent to or comparable to the ERKI method proposed by Plotz and Schmidt [16]. Following studies are planned to evaluate its performance as a directional hearing measurement tool and the reproducibility of its results.

4. Acknowledgments

This work was funded by the German Federal Ministry of Education and Research (BMBF), funding line for young engineers (13FH666IA6 and 13FH666IB6, VIWER-S).

5. Literature

1. Blauert J (1996) Spatial Hearing, Revised Ed. https://doi.org/10.7551/mitpress/6391.001.0001
2. Bregman AS (1990) Auditory Scene Analysis. https://doi.org/10.7551/mitpress/1486.001.0001
3. Bronkhorst AW (2015) The cocktail-party problem revisited: early processing and selection of multi-talker speech. Attention, Perception, Psychophys 77:1465–1487
4. Cameron S, Dillon H (2011) 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 22:678–696
5. Cameron, S., & Dillon H (2014) Remediation of spatial processing issues in central auditory processing disorder. In: Chermak Gail FM (ed) Handb. Cent. Audit. Process. Disord. Compr. Interv. Vol. 2, 2nd Ed. Plural Publishing, San Diego, CA, pp 201–224
6. Glyde H, Cameron S, Dillon H, Hickson L, Seeto M (2013) The effects of hearing impairment and aging on spatial processing. Ear Hear 34:15–28
7. Hearing Health Foundation Auditory Processing Disorders Demographic. In: Hear. Heal. Found. https://hearinghealthfoundation.org/apd-demographics. Accessed 9 May 2022
8. Iliadou V, Bamiou DE, Kaprinis S, Kandylis D, Kaprinis G (2009) Auditory Processing Disorders in children suspected of Learning Disabilities—A need for screening? Int J Pediatr Otorhinolaryngol 73:1029–1034
9. Chermak GD, Musiek FE (2013) Handbook of Central Auditory Processing Disorder, Volume II: Comprehensive Intervention, Second Edi. Plural Publishing, San Diego, CA
10. Bellis TJ, Bellis JD (2015) Central auditory processing disorders in children and adults. In: Hum. Audit. Syst. Fundam. Organ. Clin. Disord. Elsevier, Edinburgh, pp 537–556
11. Crandell CC, Smaldino JJ (2000) Classroom Acoustics for Children With Normal Hearing and With Hearing Impairment. Lang Speech Hear Serv Sch 31:362–370
12. Dockreill JE, Shield BM (2006) Acoustical barriers in classrooms: The impact of noise on performance in the classroom. Br Educ Res J 32:509–525
13. Knapp CH, Carter GC (1976) The Generalized Correlation Method for Estimation of Time Delay. IEEE Trans Acoust 24:320–327
14. Kowalk U, Doclo S, Bitzer J (2022) Signal-informed DNN-based DOA Estimation combining an External Microphone and GCC-PHAT Features. eessAS. https://doi.org/10.48550/arXiv.2206.05606
15. Maravita A, Spence C, Driver J (2003) Multisensory integration and the body schema: Close to hand and within reach. Curr Biol. https://doi.org/10.1016/S0960-9822(03)00449-4
16. Plotz K, Schmidt KK (2017) Lokalisation realer und virtueller Schallquellen mit einem automatisierten Erweiterungsmodul am Mainzer-Kindertisch – Entwicklung des ERKI-Verfahrens. Zeitschrift für Audiol 56:6–18
17. Ramírez M, Arend JM, 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: Proc. 48th DAGA. pp 282–285