forum acusticum 2026
8-12 September · Graz, Austria

ACCESSIBLE 6DOF RENDERING WITH MPEG-I IMMERSIVE AUDIO

Nils Peters[1], Andreas Silzle[2], Sascha Disch[2]

[1] *Trinity College, The University of Dublin, Ireland*
[2] *Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany*

Abstract

A growing cohort of multimedia consumers requires advanced accessibility solutions due to diverse sensory and cognitive needs such as age-related hearing loss, noise-induced hearing impairment, or congenital auditory conditions. Ensuring that emerging media technologies such as virtual and augmented reality (VR/AR) remain inclusive has become a critical requirement for broad-scale technological adoption.

Concurrently, the development of immersive Six Degrees-of-Freedom (6DoF) audio technologies represents a shift in media consumption. Unlike traditional static media playback, 6DoF audio gives listeners the freedom to navigate within a virtual acoustic scene. However, this interactivity introduces substantial challenges for the traditional authoring of accessible audio experiences during content production. The low-latency requirements of real-time 6DoF audio systems preclude complex post-processing approaches, e.g., speech enhancement in reverberant scenes containing multiple sound sources.

To enable accessible immersive audio while satisfying latency constraints, the recently finalized MPEG-I Immersive Audio standard (ISO/IEC 23090-4) integrates several accessibility tools directly within its 6DoF audio renderer. This contribution examines these integrated accessibility features within MPEG-I and demonstrates how 6DoF audio rendering can be individualized to foster more equitable experiences for individuals with diverse sensory and cognitive needs. Results from a listening test demonstrate the effectiveness of the proposed solution.

Keywords: *immersive audio, 6DoF, accessibility, hearing impairments*

Corresponding author: *nils.peters@tcd.ie.*

1. Introduction

MPEG-I Immersive Audio [1] is the new ISO/IEC standard for Six Degrees-of-Freedom (6DoF) immersive media applications, including virtual and augmented reality (VR/AR). It provides an advanced metadata framework and rendering architecture tailored towards 6DoF user requirements [2], enabling real-time interactive rendering of key acoustic effects, including distance-dependent attenuation and propagation delay, early reflections, late reverberation, occlusion, diffraction, and Doppler shifts, all based on an authored scene description and the spatial configuration (e.g., position and orientation) of sources and listener.

MPEG-I evolved from MPEG-H 3D Audio [3] to support six degrees of freedom (6DoF) interactivity, enabling listeners to freely move and orient within a three-dimensional sound field while the perceived audio scene dynamically adapts to their position and orientation [4], [5]. To compress the audio elements (objects, channels, HOA), MPEG-I uses the MPEG-H core codec [6], and can also accept PCM inputs from alternative codecs, e.g., to support future immersive communication and Social VR applications via low-latency speech codecs such as EVS [7] or IVAS [8], as demonstrated in [9].

As MPEG-I aims to deliver immersive experiences and enable Social VR interactions [10], it is vital to standardize technology that is designed to accommodate users across a wide range of hearing abilities and auditory sensitivities.

Hearing loss (HL) arises from aging, disease, ototoxic medication, genetics, and trauma, while sensory overload and auditory hypersensitivity (e.g., intolerance to loud or complex sounds) are common in neurodivergent conditions like autism. As populations age, more consumers experience some degree of HL, supporting the claim by Drever and Hugill [11] that fewer than 17% of Europeans meet ISO's definition of "normal" hearing [12].

1.1. Related Work

In recent years, the video game and AR/VR community have made significant strides toward formalizing audio accessibility [13], [14], [15], [16] and current industry standards, such as the Xbox Accessibility Guidelines [1], the W3 XR Accessibility User Requirements[2] or the IGDA Game Accessibility guidelines[3] address audio accessibility to varying degrees. For instance, it is recommended to allow separate level control of music, sound effects, voice-over, and background ambiance, so that players can mitigate cognitive overload and ensure that critical speech or screen-reader outputs remain intelligible. It is also recommended that a (non-directional) diotic headphone mix is provided to ensure that users with unilateral or strong asymmetric HL do not miss vital auditory information.

In contrast, spatial audio has also emerged as an accessibility feature. For instance, the authors of [17] indicated that binaural spatial audio can reduce auditory hypersensitivity in autistic young people. Spatial audio is also known to improve identification and localization of specific sounds in complex, crowded environments which is beneficial in many video games and AR/VR experiences. This benefit generally deteriorates with increasing HL which has been studied using hearing loss simulators [18], [19], [20].

However, while accessibility approaches for users with low vision have been well categorized and formally evaluated by the research community (e.g., [21]), tools that address audio accessibility have received considerably less research attention so far.

2. Accessibility in MPEG-I Audio

2.1. Accessibility Concept

In a 6DoF rendering environment where the listener can freely move within the scene, existing personalization strategies that presume a static listening scenario, e.g., MPEG-H's user-controllable audio group settings to increase dialog loudness [22] may not be sufficient. Novel methods, as described in the following section, account for the listener's movements and thus complement existing concepts. With MPEG-I Immersive Audio being an audio standard, we focus on accessibility strategies within the audio domain.

In MPEG-I, the 6DoF renderer composes the acoustic scene from multiple individual audio elements plus metadata placed within a simulated virtual space. This enables direct control over the scene composition. We leverage the ability to access authored metadata to directly modify the immersive real-time rendering. This avoids additional computational complexity and the risk of introducing potential artifacts (see e.g., [23], [24]) by post-processing enhancement methods such as source separation [25] or dereverberation.

[1] https://learn.microsoft.com/en-us/xbox/accessibility/xbox-accessibility-guidelines/105

[2] https://www.w3.org/TR/xaur

[3] https://igda-gasig.org/get-involved/sig-initiatives/resources-for-game-developers/sig-guidelines

2.2. Accessibility Interface

As illustrated in Fig. 1, the modification of parameters that control the rendering is facilitated by a decoder-side Accessibility Interface, which maps high-level user-specific accessibility user parameters to low-level rendering parameters via a Data Preprocessor. This preprocessor accounts for the user's spatial pose and position within the scene, as well as authoring constraints embedded in the bitstream (e.g., indicators specifying audio elements that are exempt from modification). The preprocessing operates at the metadata (control data) level and does not involve real-time audio rendering. Decoding and rendering are performed by the Decoder and Renderer functional block, which processes the incoming MPEG-I bitstream and synthesizes the sound field in accordance with the metadata and the user's spatial context. Because our approach only modifies the low-level rendering instructions, there is no increase in the system's low motion-to-sound latency [5] or a significant increase in run-time complexity.

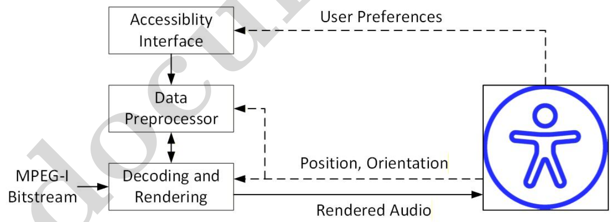

Figure 1: Architecture to configure accessible rendering.

2.3. Accessibility Tools

The accessibility tools are organized into categories as described in the following section. The goal is to let users effectively adapt the audio rendering to their individual needs with minimal effort.

Notably, the MPEG-I standard defines the Accessibility Interface and its input user parameters but does not prescribe predefined configurations, as accessibility settings are inherently context and user-dependent. Hence, this design enables personalized configurations, e.g., by audiologists or third-party algorithms, which lie outside the scope of the standard, while preserving flexibility for future enhancements and optimization strategies.

2.3.1. Attenuation of Early Reflections and Late Reverberation

Early reflections and especially late reverberation significantly alter clarity and speech intelligibility [26], [27]. The accessibility user parameters provide control to independently attenuate the energy of early reflections and late reverberation in the audio scene.

2.3.2. Expansion/Compression of Distance-dependent Attenuation

A primary cue for distance perception is distance-based sound attenuation that is often simulated in virtual auditory environments using the $1/r$ free-field approximation: doubling the distance r causes an attenuation of 6 dB SPL. Distant sources sound quieter, and sources that are in close proximity sound loud. An accessibility user parameter provides a control of the tuning parameter α, to artificially expand or compress the distance used for the distance-dependent gain attenuation by replacing the distance r with r^{α}. The intended effect is that the depth of a sound scene, comprised of audio objects, can be expanded or compressed by the user at runtime. Fig. 2 plots the distance-dependent gain attenuation for three configurations, including the default setting of the renderer (red curve). For instance, when $\alpha > 1$ the depth of the scene expands and objects that are farther away will become less audible and vanish (blue curve), making the sound scene less complex and easier to attend to nearby sounds while navigating the virtual world.

In contrast, when $\alpha < 1$ (yellow curve), the depth of the scene shrinks and objects that are farther away will become more audible.

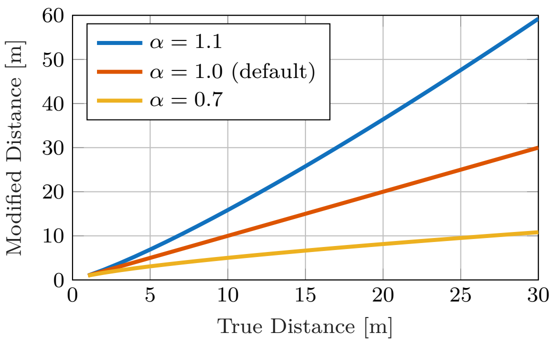

Figure 2: Effect of the parameter α on the distance r used to render distance-dependent attenuation.

2.3.3. Directional Focus

In complex audio scenes, e.g., many simultaneous talkers, it can be helpful to reduce the scene complexity by attenuating sounds that are outside the area of interest, which is often centered around the user's viewing direction. For this use case, the Accessibility user parameter can configure a directional focus effect which attenuates sound energy outside of the direction of interest. As depicted in Fig. 3, the aperture width, the amount of attenuation outside the direction of interest, and the transition width between these two regions are configurable. By default, this directional filter aligns with the user's forward direction, but can be steered in real-time to other directions, e.g., using a handheld controller or an eye-tracker.

2.3.4. Spectral Compensation

These parameters describe the magnitude spectrum to equalize the binauralized sound field for the left and right ear independently. Based on this magnitude spectrum, minimum-phase filters are designed at the initialization phase and applied during binauralization without additional block delay.

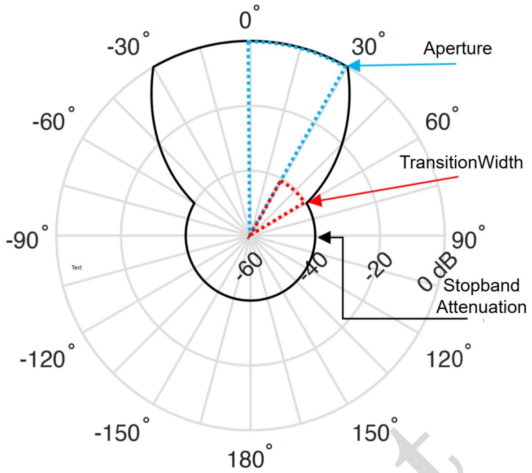

Figure 3: Visualization of the directional focus beam pattern and its configuration parameters.

3. Subjective Evaluation

3.1. Setup

To evaluate the potential of the described features, a VR listening test was conducted. We extended the Audio Evaluation Platform (AEP) [28] that was designed for the development process of MPEG-I, including verification testing [29].

To approximate how the rendered MPEG-I scenes sound to hearing-impaired listeners, we integrated the Hearing Loss Simulator (HLS) VST plugin by Mourgela et al. [20] which approximates the four common hearing-loss effects:

- High-frequency attenuation
- Spectral smearing
- Temporal disruption
- Rapid loudness growth

For more details of these hearing-loss effects and hearing loss simulators, see e.g., [20], [30]. In the evaluation setup the HLS plugin processes the binaural output of the MPEG-I immersive audio renderer in real-time (see Fig. 4). This allows participants with normal hearing abilities to compare and evaluate the proposed accessibility features using the MPEG-I testing methodology while being exposed to various types of simulated hearing loss.

3.2. Test Conditions

The listening test consists of four VR test scenes. All of them are primarily composed of audio objects:

Scene A: Sonically complex scene with ambient battle noise, tank and helicopter noise and a jet flyover. A nearby radio emits distorted speech.

Scene B: Highly reverberant scene of a large cathedral with four singers in front of the altar.

Scene C: Moderately reverberant gymnasium with three simultaneous talkers (see Fig 5).

Scene D: A large outdoor scene, including a reverberant indoor swimming pool with a conversation

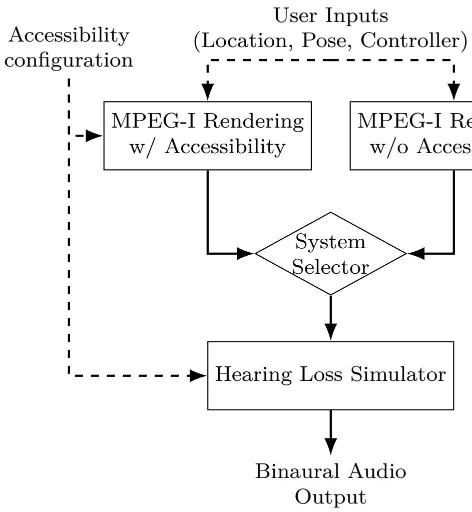

Figure 4: Listening test setup using the Audio Evaluation Platform (AEP) and the Hearing Loss Simulator (HLS).

between two swimmers. Another swimmer making splashing noises.

Figure 5: Scene C: Three simultaneous talkers in a gym.

For each scene, three different simulated hearing loss conditions are tested, ranging from a) mild HL in both ears; b) moderate HL in both ears; to c) mild HL in left ear and moderate HL in right ear (see details in Tab. 1 and Fig. 6), resulting in a total of 12 test trials. In each trial, two MPEG-I rendering conditions are compared: Without Accessibility and With Accessibility. Since our focus is on improving listening effort and speech intelligibility, all four VR scenes include voice or speech content, along with environmental sound effects, and comprise indoor and outdoor environments with corresponding acoustic rendering of early reflections and reverberation. As an example, Fig. 5 shows Scene C with three simultaneous talkers in a reverberant gym, simulating the challenging *Cocktail Party* situation where a listener has to concentrate on one particular speaker in a noisy crowd. For each trial, an accessibility configuration, prepared during a pilot study, was loaded as shown in Fig. 4, aiming to balance enhanced clarity with preserving the intended immersion of the scene.

Table 1: Settings to simulate Mild and Moderate hearing loss using the HLS VST plugin [20].

Hearing-Loss Effect	Mild HL	Moderate HL
High-freq. Attenuation	mild	moderate
Spectral Smearing	low	low
Rapid Growth	on	on
Temporal Disruption	off	on

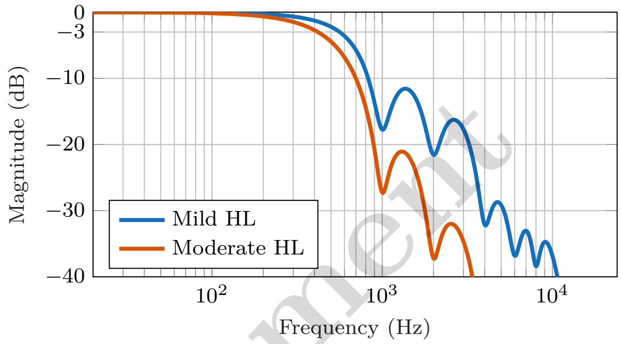

Figure 6: High-frequency loss effect as simulated by the HLS for the Mild HL and Moderate HL setting.

3.3. Subjects and Listening Task

Eight normal-hearing (NH) participants, all with prior experience in VR listening tests and the AEP, took part in the evaluation. Using a double-blind A-B testing methodology, participants were asked to "*Compare your Listening Effort and the Speech Intelligibility between the two systems under test.*". Using a HTC Vive Pro HMD and Beyerdynamic DT990 Pro headphones, subjects could freely move within the VR scenes. Using the hand-held controller, a popup GUI within the VR scene allowed users to switch between the two rendering conditions continuously at runtime and to provide their rating on a 7-point bipolar scale that ranges from "A is much better" to "B is much better" (see Fig. 7).

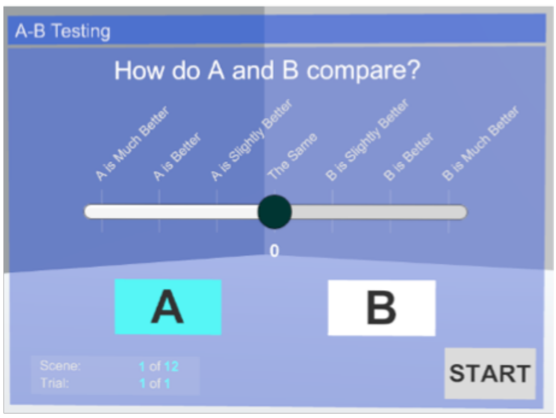

Figure 7: Evaluation GUI displayed within the VR scene.

3.4. Results

Fig. 8 presents the subjective data of the A-B listening test with the three previously defined hearing loss configurations and four VR scenes as boxplots. On average, the results show a consistent preference for the accessibility condition across scenes and simulated HL configurations.

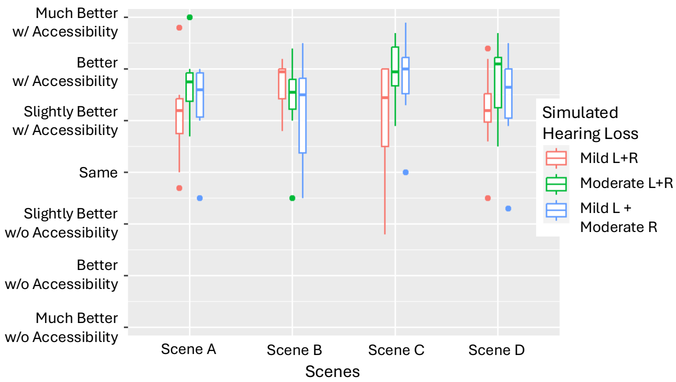

Figure 8: Listening test results of A-B test for the three different hearing loss simulations.

Fig. 9 shows the pooled results across all four scenes, indicating an improvement between “Slightly Better” and “Better” for each of the simulated HL conditions. The highest improvement was achieved for the moderate HL condition.

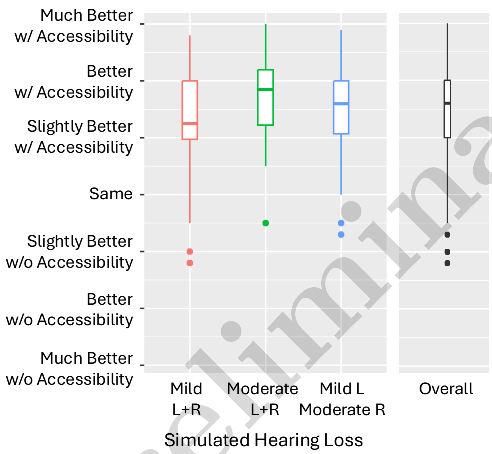

Figure 9: Pooled results across scenes for the three different hearing loss simulations and overall.

Outliers below and above the median values are present in many conditions. For Scene C and the Mild HL, scores show the greatest variability, suggesting varying preferences across listeners. This indicates that (generic), pre-defined accessibility configurations may have limited benefit. Also, during a post-test debriefing, some participants reported that the tested accessibility configurations generally improved their perception, but the improvements varied in larger scenes that feature different acoustic environments. This indicates that in practice, users will benefit from the ability to quickly bypass or modify the accessibility user settings.

4. Conclusion and Summary

In this contribution, we reported the motivation, concepts, and implementation of the audio accessibility options within the MPEG-I immersive audio standard. Being a fully interactive 6DoF renderer, MPEG-I enables advanced possibilities for powerful accessibility features to address various forms of Hearing Loss (HL) compared to existing approaches for 3DoF audio systems. We discuss the technical means that implement these features and demonstrate how 6DoF audio rendering can provide customized experiences adapted to diverse sensory and cognitive needs without any increase in motion-to-sound latency or significant complexity overhead.
An A-B listening test was conducted in which three HL configurations were simulated in four different VR experiences. Overall, on a bipolar 7-point scale, an improvement between 1 and almost 2 points for each HL condition was achieved. The listening-test results, obtained with simulated hearing-loss conditions, indicate that MPEG-I Immersive Audio has the potential to support listeners with a range of hearing abilities. Especially in social-VR contexts, social participation and inclusion are key requirements that MPEG-I Immersive Audio addresses by design.

5. Ethics Statement

All participants provided their informed consent prior to the listening test and agreed to the reporting of anonymized data.

References

[1] ISO/IEC, *23090-4, information technology — coded representation of immersive media, part 4: MPEG-I immersive audio*, Geneva, Switzerland, 2025.

[2] S. R. Quackenbush and J. Herre, “MPEG standards for compressed representation of immersive audio,” *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1578–1589, 2021. DOI: 10.1109/JPROC.2021.3075390.

[3] ISO/IEC, *23008-3, information technology — high efficiency coding and media delivery in heterogeneous environments — part 3: 3D Audio*, 2019.

[4] J. Herre, S. Disch, C. Borß, A. Silzle, A. Adami, and N. Peters, “MPEG-I immersive audio: A versatile and efficient representation of VR/AR audio beyond point source rendering,” in *AES 2024 International Audio for Games Conference*, Tokyo, Japan, 2024.

[5] S. Disch et al., “MPEG-I Immersive Audio - The ISO/MPEG Standard for Virtual / Augmented Reality Audio,” *Journal of the Audio Engineering Society*, vol. 74, 2026. DOI: 10.17743/jaes.2022.0276.

[6] R. L. Bleidt et al., "Development of the MPEG-H TV audio system for ATSC 3.0," *IEEE Transactions on Broadcasting*, vol. 63, no. 1, pp. 202–236, 2017. DOI: `10.1109/TBC.2017.2661258`.

[7] 3GPP, *Ts 26.441: Codec for enhanced voice services (EVS); general overview*, 2016.

[8] 3GPP, *Ts 26.250: Codec for immersive voice and audio services (IVAS); general overview*, 2025.

[9] N. Peters et al., "Demonstrating interoperable immersive audio communication for the metaverse," en, in *IEEE 5th International Symposium on the Internet of Sounds (IS2)*, Erlangen, Germany, 2024.

[10] A. Silzle et al., "Perceptual evaluation of own voice and social virtual reality using the MPEG-I immersive audio standard," in *159th AES Convention*, Long Beach, CA, USA, 2025.

[11] J. L. Drever and A. Hugill, *Aural Diversity.* Routledge, 2022.

[12] ISO, *226, acoustics — normal equal-loudness-level contours*, Geneva, Switzerland, 2023.

[13] S. Serafin, A. Adjorlu, and L. M. Percy-Smith, "A review of virtual reality for individuals with hearing impairments," *Multimodal Technologies and Interaction*, vol. 7, no. 4, p. 36, 2023.

[14] F. Heilemann, G. Zimmermann, and P. Münster, "Accessibility guidelines for VR games - a comparison and synthesis of a comprehensive set," *Frontiers in Virtual Reality*, vol. 2, p. 697 504, 2021.

[15] A. Borna, S. Z. Mousavi, F. Fathollahzadeh, A. Nazeri, and R. E. Harari, "Applications of augmented and virtual reality in enhancing communication for individuals who are hard of hearing: A systematic review," *American Journal of Audiology*, vol. 33, no. 4, pp. 1378–1394, 2024.

[16] J. Dudley, L. Yin, V. Garaj, and P. O. Kristensson, "Inclusive immersion: A review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse," *Virtual Reality*, vol. 27, no. 4, 2023. DOI: `10.1007/s10055-023-00850-8`.

[17] D. Johnston, G. Kearney, and H. Egermann, "The use of binaural based spatial audio in the reduction of auditory hypersensitivity in autistic young people," *International journal of environmental research and public health*, 2022.

[18] C. Valzolgher, S. Capra, K. Sum, L. Finos, F. Pavani, and L. Picinali, "Spatial hearing training in virtual reality with simulated asymmetric hearing loss," *Scientific Reports*, vol. 14, no. 1, 2024.

[19] M. Cuevas-Rodriguez et al., "An open-source audio renderer for 3D audio with hearing loss and hearing aid simulations," in *142nd AES Convention*, Berlin, Germany, 2017.

[20] A. Mourgela, J. Reiss, and T. R. Agus, "Investigation of a real-time hearing loss simulation for use in audio production," in *149th AES Convention*, 2020.

[21] Y. Zhao, E. Cutrell, C. Holz, M. R. Morris, E. Ofek, and A. D. Wilson, "Seeingvr: A set of tools to make virtual reality more accessible to people with low vision," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–14.

[22] C. Simon, M. Torcoli, and J. Paulus, *MPEG-H audio for improving accessibility in broadcasting and streaming*, 2019. DOI: `0.48550/arXiv.1909.11549`. arXiv: `1909.11549 [eess.AS]`.

[23] J. Agnew, "The causes and effects of distortion and internal noise in hearing aids," *Trends in amplification*, vol. 3, no. 3, pp. 82–118, 1998.

[24] S. M. K. Madsen and B. C. J. Moore, "Music and hearing aids," *Trends in Hearing*, vol. 18, 2014, PMID: 25361601. DOI: `10.1177/2331216514558271`.

[25] M. Sunohara, C. Haruta, and N. Ono, "Low-latency real-time blind source separation for hearing aids based on time-domain implementation of online independent vector analysis with truncation of non-causal components," in *IEEE International conference on acoustics, speech and signal processing (ICASSP)*, 2017.

[26] D. Fogerty, A. Alghamdi, and W.-Y. Chan, "The effect of simulated room acoustic parameters on the intelligibility and perceived reverberation of monosyllabic words and sentences," *The Journal of the Acoustical Society of America*, vol. 147, no. 5, 2020.

[27] G. R. Popelka, B. C. Moore, R. R. Fay, and A. N. Popper, *Hearing aids.* Springer, 2016, vol. 56.

[28] T. Robotham, O. Rummukainen, J. Herre, and E. A. Habets, "Evaluation of binaural renderers in virtual reality environments: Platform and examples," in *145th AES Convention*, New York, USA, 2018.

[29] A. Silzle et al., "Perceptual evaluation of the MPEG-I Immersive Audio standard," in *160th AES Convention*, Copenhagen, Denmark, 2026. [Online]. Available: `https://aes.org/publications/elibrary-page/?id=23149`.

[30] T. Deutsch and J. Fels, "Evaluating models for the simulation of cochlear hearing loss," in *Proc. of DAGA*, Hamburg, Germany, 2023.