Politecnico di Torino
Doctoral Dissertation
Doctoral Program in Bioengineering and Medical-Surgical Sciences (37.th cycle)

Set-up and investigation of a virtual reality system for clinical usage aimed at more ecological hearing assessments

Angela Guastamacchia

Supervisors

Prof. Arianna Astolfi, Supervisor
Prof. Fabrizio Riente, Co-supervisor
Ph.D. Giuseppina Emma Puglisi, Co-supervisor

Summary

According to the World Health Organization, over 700 million people are expected to suffer from disabling hearing loss by 2050, significantly affecting their life quality, hindering speech communication and leading to social isolation. Although modern Hearing Aids (HAs) can enhance auditory perception, their effectiveness in real-world multi-talker, noisy, dynamic environments with adverse acoustics remains limited. One key reason lies in the limitations of current clinical protocols. Standard laboratory tests focus primarily on pure-tone audiometry, with fewer than 20% of clinicians incorporating speech audiometry, which nevertheless lacks ecological validity—that is, the outcomes fail to reflect real-life hearing-related functionality. Speech intelligibility (SI) tests typically use isolated words or sentences in quiet or with stationary noise, presented within acoustically treated booths via simple loudspeaker setups, thus neglecting the complexity of real-life auditory scenes and omitting Self-Motion (SM) and visual cues essential for localization and speech comprehension. While in-field testing offers ecological validity—with patients reporting their real-life experience to the audiologist—, its lack of control over test conditions undermines HA fine-tuning accuracy. To bridge this gap, hearing research has turned to Virtual Reality (VR), which enables controlled reproduction of immersive AudioVisual (AV) scenarios emulating real-life conversation conditions in the lab. However, most systems in research contexts often rely on expensive setups for accurate sound field reproduction—requiring dozens of loudspeakers, significant computing resources for AV simulations, software expertise, and facilities—making them impractical for clinical use.

To promote the integration of more ecologically valid SI tests into standard clinical practice protocols, this thesis introduces and evaluates, in terms of its clinical applicability, a cost- and complexity-efficient VR system that leverages well-established spatial audio rendering techniques and can be easily operated in clinical settings by non-expert users. The system features a spherical 16-speaker array using 3rd-Order Ambisonics (3OA) rendering, synchronized with a VR headset to display a 360° visual scene. To support intuitive clinical use, a preliminary and easily expandable database of immersive communication scenarios based

on real-world AV recordings has been assembled, from which clinicians can draw to set up the AV scene in which to place the SI test through a user-friendly, purpose-built application that also allows to store patient histories, run AV tests, and record results. The system investigation is approached from multiple perspectives.

A preliminary validation comparing a real reverberant room with its VR reproduction—based on both physical assessments (monaural and binaural room acoustic parameter measurements) and a perceptual SI test on subjects across five auditory scenes—reveals no statistical differences in perceived SI, with most acoustic parameters within the just noticeable difference. Subjective evaluations through questionnaires on the overall AV experience further confirm the VR system elicits a satisfactory sense of presence within the immersive scenes.

As a further step, an inter-laboratory comparison involving different listening rooms and both Italian and German participants is presented. The proposed VR system—i.e. (i) a 3D 3OA 16-speaker setup tested with Italian participants—is compared against a more complex setup—(ii) a 3D 7OA 45-speaker setup—and two other 2D 16-speaker setups tested with German participants—(iii) a 2D 7OA setup with equally spaced speakers, and (iv) a 2D Vector Base Amplitude Panning setup with higher frontal speaker density—based on four perceptual tests relevant for clinical applications: sound localization, Minimum Audible Angle (MAA), spatial audio quality, and SI. Results revealed no significant differences among the setups in MAA and perceptual quality ratings. Lower localization errors are observed for setup (iv) in the azimuth plane and for setups (i) and (ii) performing equally in the elevation plane. Different SI results are found across setups for the most complex auditory scenes, with setup (i) generally yielding worse SI in the most reverberant condition, and setup (iv) performing equally or better than the others. These differences may stem from a blurred sound image in 3OA reproduction or language-specific interactions with scene complexity, highlighting the need for further research to check cross-linguistic comparability of validated speech materials in ecological scenarios.

Finally, a systematic investigation of the combined effects on SI of non-auditory aspects in real-life communication scenarios—such as SM, contextual, sound source position, and target talker lip-sync-related visual cues—is presented, aiming to identify which factors are truly worth integrating into the proposed VR system to enable more ecological testing. SI test results for different AV scenes set in a reverberant conference hall, conducted with participants within the VR system, confirm that lip-sync-related visual cues strongly supports SI—especially when real video recordings rather than synthetic avatars are used. Conversely, the other visual cues alone do not enhance SI and particularly SM leads to the poorest SI outcome, possibly due to participants' distraction.

Overall, the proposed VR setup proves suitable for clinical use, particularly when elevation perception needs to be assessed, supporting more ecological SI testing and advancements in clinical hearing evaluation protocols.