AURALYS: Smart Glasses to Improve Audio Selection and Perception in Educational and Working Contexts

Gianluca Filippini[1], Guido Borghi[2,*], Enrico Giliberti[2,*], Paola Damiani[2,*] and Roberto Vezzani[1]

[1]*Dipartimento di Ingegneria "Enzo Ferrari", University of Modena and Reggio Emilia*
[2]*Dipartimento di Educazione e Scienze Umane, University of Modena and Reggio Emilia*

Abstract

The ability to discern multiple sound sources in complex environments is an innate auditory skill that varies across individuals due to diverse personal and contextual factors. Conditions such as aging, disabilities, or neurodevelopmental disorders – now more widely recognized – highlight the need for inclusive approaches. Hearing impairment is commonly understood as deafness or hearing loss, but numerous conditions affect not the quantity (how much one hears) but the quality of auditory perception (how one hears). This calls for interdisciplinary research on how technological and AI tools can support diverse users, promoting inclusion and improving quality of life, particularly for those with vulnerabilities. Therefore, in this paper, we introduce and discuss the adoption of AURALYS, smart glasses expressively designed to improve audio capabilities in educational and working scenarios. In particular, this device is intended to enhance audio selection and perception in dynamic contexts, in which multiple competing voices and background noises are present. We also introduce the VERSE framework to create and collect synthetic audio data to train machine learning systems for audio selection and perception implemented on the smart glasses.

Keywords

Audio Capability, Selective Hearing, Smart Glasses, Artificial Intelligence

1. Introduction

The ability to perceive and distinguish multiple sound sources in complex acoustic environments – recognized as an innate auditory skill, akin to other cognitive processes – manifests differently among individuals depending on various endogenous and exogenous factors, leading to diverse profiles of competence and functioning. Some of these conditions are universal, such as aging, while others stem from specific individual circumstances, such as sensory disabilities or neurodevelopmental disorders. These conditions are increasingly present in educational and work settings, due both to greater recognition and the spread of an inclusive culture.

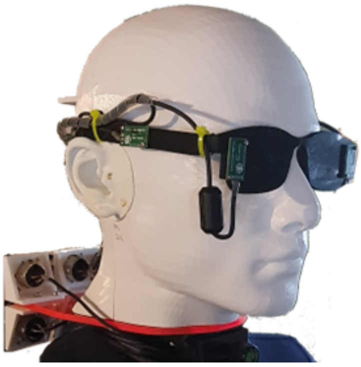

Figure 1: The prototype of AURALYS glasses placed on a 3D printed head.

It is therefore essential to foster interdisciplinary reflection closely tied to the purposes of using technological devices and to the characteristics of individuals and contexts. This should be done through a multidisciplinary approach, starting from a pedagogical perspective that values the emancipatory potential of technology and AI in enhancing quality of life for all, with those who experience vulnerability.

2nd International Workshop on AI in Society, Education and Educational Research (AISEER)
*Corresponding author.

gianluca.filippini@unimore.it (G. Filippini); guido.borghi@unimore.it (G. Borghi); enrico.giliberti@unimore.it (E. Giliberti); paola.damiani@unimore.it (P. Damiani); roberto.vezzani@unimore.it (R. Vezzani)
https://aimagelab.ing.unimore.it/ (G. Filippini); https://aimagelab.ing.unimore.it/ (G. Borghi); https://aimagelab.ing.unimore.it/ (R. Vezzani)
0000-0003-2441-7524 (G. Borghi); 0000-0002-9790-617X (P. Damiani); 0000-0002-1046-6870 (R. Vezzani)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2025-12-18

For example, in schools, students spend many hours in noisy environments that can create challenging acoustic conditions, with background noise and overlapping voices interfering with lesson comprehension and the ability to distinguish individual speakers [1] (*e.g.*, crowded classrooms, gyms, hallways, cafeterias). Similar conditions can be found in workplace settings, such as offices. Industrial environments – such as factories and production plants – pose additional challenges, where the use of loud machinery requires appropriate hearing protection. However, such protective equipment may also reduce auditory sensitivity or impair the ability to discriminate sounds and identify their sources.

Considering these elements, we introduce and discuss a technological solution called AURALYS, *i.e.*, smart glasses provided with embedded audio capturing and analysis capabilities (see Fig. 1). From a hardware point of view, the proposed system is mainly made up of six microphones, placed on a frame that, in future developments, could also integrate cameras and other sensors. The number and positioning of microphones is itself an object of study and research to maximize the ability to localize audio sources and, at the same time, limit the computational capabilities required for the corresponding processing. AURALYS also integrates several software components capable of processing the signals coming from the microphones in real time. Most of them are realized through innovative machine learning techniques, which exhibit excellent performance but, at the same time, require large amounts of data to obtain valid training. With current technologies, the ability of the system to generalize to any condition and situation is incompatible with a low-latency embedded system suitable for integration in AURALYS. Therefore, it is necessary to create specific datasets that contain only the degrees of freedom strictly necessary for the application in question. For this purpose, the VERSE framework was introduced, a complete framework able to generate suitable datasets of synthetic but realistic recordings of human voices together with all the annotations required to train machine learning algorithms. The specific dataset can be configured to include, for example, specific languages, types of voice, mutual positioning, and motion between audio sources and the listener.

2. Related Work

2.1. Inclusive Perspectives of Audio Capability

Recent studies have investigated the impact of hearing loss on other functions essential for living a quality life, confirming that it can significantly impair learning processes, affecting language development [2] and broader language skills [3]. These impacts potentially affect a wide range of individuals across the lifespan and in various settings, from schools to workplaces and elderly care environments.

Hearing impairment is commonly understood as deafness or hearing loss, but numerous conditions affect not the quantity (*how much* one hears) but the quality of auditory perception (*how* one hears). As noted by Bérard [4], these include auditory slowness, painful hearing, lack of auditory selectivity, auditory dis-laterality, auditory distortions, residual hearing effects, and tinnitus. All of these anomalies can significantly impact attention and learning. During developmental age, beyond more evident conditions such as deafness or hearing loss, the ability to focus selectively can also be compromised in cases of neurodevelopmental disorders, such as Attention Deficit Hyperactivity Disorder (ADHD) and Specific Learning Disabilities (SLDs) – including dyslexia, dysgraphia, and dyscalculia – and Autism Spectrum Disorder. Research has highlighted the key role of Executive Functions [5] and the difficulties associated with deficits in attentional and perceptual abilities [6].

Although auditory processes have been studied less frequently than visual ones, impairments in auditory attention have been shown to be significant in students with ADHD, SLDs, and Disruptive Behavior Disorders. Low scores in auditory attention are associated with reduced selective and sustained attention [7], leading to notable consequences for both learning quality and active participation.

2.2. Selective Hearing

Selective Hearing [8] becomes less effective with increasing complexity of the environment, which means a larger number of competing sound sources and a higher background noise level. A key aspect of this perceptual process involves localizing where sounds are coming from, which the human brain achieves by using a combination of spatial, spectral, and temporal cues. Among the most critical spatial cues are Interaural Time Differences (ITD) and Interaural Level Differences (ILD). ITD refers to the tiny differences in the time it takes for a sound to reach each ear; for example, if a sound source is located to the left of a listener, it will reach the left ear slightly earlier than the right. This time difference, often in the range of microseconds, is processed by the auditory system to infer the direction of the sound in the horizontal plane. ILD, on the other hand, refers to differences in sound pressure level (or loudness) between the ears, which occur because the head acts as a physical barrier, casting an acoustic shadow that attenuates sounds arriving at the far ear. These interaural cues are most effective for high-frequency sounds (ILD) and low-frequency sounds (ITD), respectively, and are combined by the brain to localize sound sources with remarkable precision.

In near-field scenarios, where sound sources are located close to the listener, the auditory system can also exploit additional spatial cues, such as variations in the shape and timing of reverberations and subtle changes in binaural cues due to head movement. Moreover, proximity of the source often increases the signal-to-noise ratio and preserves finer acoustic details, making it easier to distinguish between individual voices. In contrast, far-field conditions introduce challenges such as increased reverberation, reduced spatial separation, and signal degradation due to distance, all of which blur spatial and spectral distinctions between sources. Auditory spectral differences become predominant for subjects with hearing aid devices. Differences in spectrum perception between the two ears will affect the acoustic signal arriving at the two eardrums.

2.3. Technical Background for Selective Hearing

Humans are able to localize audio sources as a combination of multiple senses. Audio cues are the fundamental part of this process, even if it is proven that the interaction with visual information enhance the capability to distinguish sounds sources [9]. For this reason, the usage of multi-microphone techniques have raised the interest of researchers, exploring techniques like beamforming to improve source localization in combination with head orientation and gaze [10]. Arrays of microphones have been used to collect data for hearing aid applications, opening new scenarios [11].

In recent years, deep learning-based models have significantly advanced the state-of-the-art in both sound source separation and localization. This is also the case for more complex scenarios involving hearing aid implants [12]. However, the fields face persistent challenges that hinder systematic progress and fair benchmarking, primarily related to open-source dataset availability and reproducibility of results. Despite the availability of reference benchmarks like CHiME [13] and DCASE [14], it is still difficult to obtain the same results starting from a single product scenario, facing differences on microphones, geometry, and calibration. Reference dataset do not fully capture the complexities of real-world acoustic environments, such as custom reverberation, dynamic source movement, background noise, and overlapping speech from multiple direction; all applied to a real, specific receiver (human, binaural or multi-microphone array) that is different from the one used the reference dataset.

To mitigate the effort required for acquiring huge set of real recording, despite the complexity of measurements and data processing, the usage of synthetic data in combination with direct audio recordings presents advantages in scalability and reproducibility, allowing for solving some challenges presented by "fixed recording audio datasets". Datasets with accurate spatial annotations of all the components of the audio chain (*e.g.*, microphone array geometries, source coordinates in space, sound levels with calibration) became available only in recent times, limiting the capability to train or evaluate models that rely on spatial cues for localization.

However, even when datasets are available, there is a lack of consistency in evaluation protocols, metrics, and data splits. This leads to a reproducibility gap in which results reported in the studies cannot be directly compared. Furthermore, some datasets used in high-profile publications are not made publicly available due to licensing restrictions or privacy concerns, making it challenging for researchers to validate or extend previous work. Finally, when considering the usage of synthetic data, it is important to address the reverberation of the environment to properly simulate the audio signal as close as possible to the real-life scenario. The shape and size of the environment surrounding sound sources influence early reverberations, which are predominant in the source localization process [15].

2.4. Synthetic Audio Data Generation

The AURALYS glasses and the developed software rely on the assumption that synthetic datasets can be created and correctly used to train deep learning algorithms. In particular, one of the most important is the generation of realistic asset items to combine in the rendering framework, and the study of the intrinsic characteristics of the listener is one of them.

The scientific study of binaural hearing has its roots in psychoacoustics and auditory physiology, dating back to the early 20th century [16]. By the mid-20th century, advances in signal processing and acoustics enabled more rigorous measurement of Head-Related Transfer Functions (HRTFs), a critical component in binaural modeling. Indeed, HRTFs characterize how the listener's head, torso, and pinnae filter incoming sound before reaching the eardrums, and provide direction-dependent spectral and temporal cues essential for precise localization [17]. With improvements in audio recording equipment and microphones, and thanks to digital signal processing applied to audio signals, it has been possible to define multiple techniques to measure the HRTF function of a given subject [18]. Different techniques have different performance regarding signal-to-noise ratio (SNR) and spatial accuracy [19], but the Exponential Sine Sweep stands out as one of the most used techniques up to modern times [20].

Physical dependencies of the HRTF from human body characteristics have been studied with measurements for ear pinna, torso, and head, forming databases of morphological measurements. Yet the collection of these measurements requires careful setup and calibration to properly retrieve the transfer function. Measuring HRTFs is still a time consuming and complex procedure, often requiring anechoic chambers. New techniques have been proposed to simplify constraints on the recording environment [21].

HRTF measurements are even more important when we focus on human voice and speech intelligibility. Using non individualized HRTFs will introduce a significant difference between the reference data and the real life scenario. This is critical for synthetic datasets, which are widely used for datasets related to neural network development [22].

3. AURALYS: smart glasses with AUdio captuRing and anALYSis capabilities

AURALYS [1] represents a cutting-edge research project that aims to design a tool to enhance human auditory perception in dynamic, real-world environments. The idea behind the adoption of glasses to improve audio capabilities derives from he necessity to include an array of microphones, in which each devices is slightly distanced from the others. In our opinion, the temples of glasses represent a good solution to place this microphone array, due to the physical space available on a rigid surface, and the proximity to the ears. In addition, the use of glasses is largely accepted in society, especially in contexts related to school and work. We also note that, as future work, there is the possibility of easily expanding the functionalities of glasses through the use of vision systems in terms of, for instance, two cameras placed on the lenses.

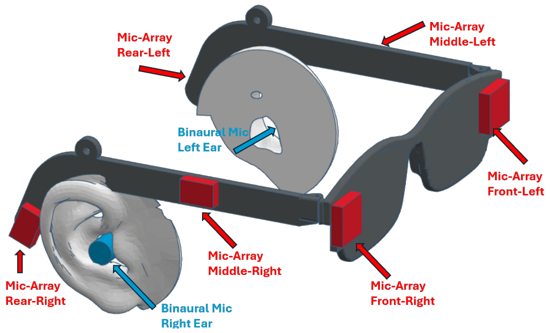

Figure 2: The 3D rendering of the AURALYS glasses; highlighted in red, the position of the microphones.

[1] https://github.com/iot-unimore/Auralys

AURALYS integrate a custom-designed microphone array mounted on 3D-printed frames (see Fig. 2), precisely positioned to capture spatial audio cues around the wearer. Unlike standard binaural recording setups, the glasses leverage six analog microphones – strategically placed to maximize directional sensitivity – enabling advanced real-time audio processing, including source localization, selective hearing, and speech enhancement. The six microphones are not only spatially distributed along the transverse (horizontal) plane, but also have a difference in positioning on the craniocaudal axis (distance from the ground), thus enabling a more correct localization of the audio sources in the entire three-dimensional space.

Thanks to the use of analog microphones and high-fidelity acquisition systems, the glasses achieve tight synchronization between input channels, which is essential for computing accurate HRTFs (see Sect. 2.4). These functions capture how sound interacts with the unique geometry of the user's head and ears, ensuring highly individualized and realistic spatial audio rendering. This makes the AURALYS glasses particularly effective in acoustically challenging scenarios, such as environments with reverberation, overlapping voices, or moving sound sources.

By combining hardware precision with the VERSE software framework – which simulates and processes dynamic scenes with high realism – AURALYS is not only a platform for research but also a promising assistive device. Indeed, as mentioned, they open up new possibilities for augmented hearing, enhanced situational awareness, and robust speech understanding in settings like crowded streets, classrooms, or public transportation. In this way, AURALYS stands at the intersection of wearable technology, human-centered design, and advanced acoustic modeling.

3.1. Hardware Prototype

The 3D printed glasses prototype is the base to place the six microphones in specific positions (see Fig. 2). Analog mems microphones are used to simplify the synchronization of recorded signals with the source stimuli. As shown in Figure 1, glasses are placed on a 3D printed head, in order to replicate a realistic setting during the sound acquisition. Industry does provide standardized mannequins replicating the human torso and head. Few products are available on the market like the well known Kemar mannequin[2]. These apparatuses are a must to have in acoustic and physic research, but there are use cases where the subject is custom and it is not comparable to the standard specification.

In our work, we use a 3D printed head from the open-source project OpenAural [23], available under common-creative license[3]. The OpenAural head has been selected for its license and reproducibility at low cost, but with modern tools it is possible to perform a 3D scan and print of any subject, robotic device or human and, in particular, child head. For this project the printed head is combined with a commercial torso mannequin, similar to the ones used for store display. The receivers are built using an analog mems microphone model KNOWLES SPM0687LR5H-1. The full schematic is released as part of the repository, for reproducibility, and provide a small microphone with 48 volts phantom power capabilities.

The technique used to compute the HRTF function is based on the common sine sweep method, where the stimuli is produced by a calibrated (equalized) speaker and the receivers and source signal are recorded with a digital audio card on a computer. AURALYS project uses a FAITAL PRO audio speaker 4FE32 (8 ohms)[4]. The combination of speaker and audio amplifier has been equalized for a flat audio response via external equalizer Beheringer UltraCurve DEQ2496[5]

3.2. Software modules

The software developed for the project and capable of processing the audio streams coming from the six microphones of the glasses must include the components shown in Figure 3. For each specific application,

[2] https://www.grasacoustics.com/industries/audiology/kemar

[3] https://www.thingiverse.com/thing:4691843

[4] https://faitalpro.com/it/products/LF_Loudspeakers/product_details/index.php?id=401005100

[5] https://www.behringer.com/product.html?modelCode=0821-AAD

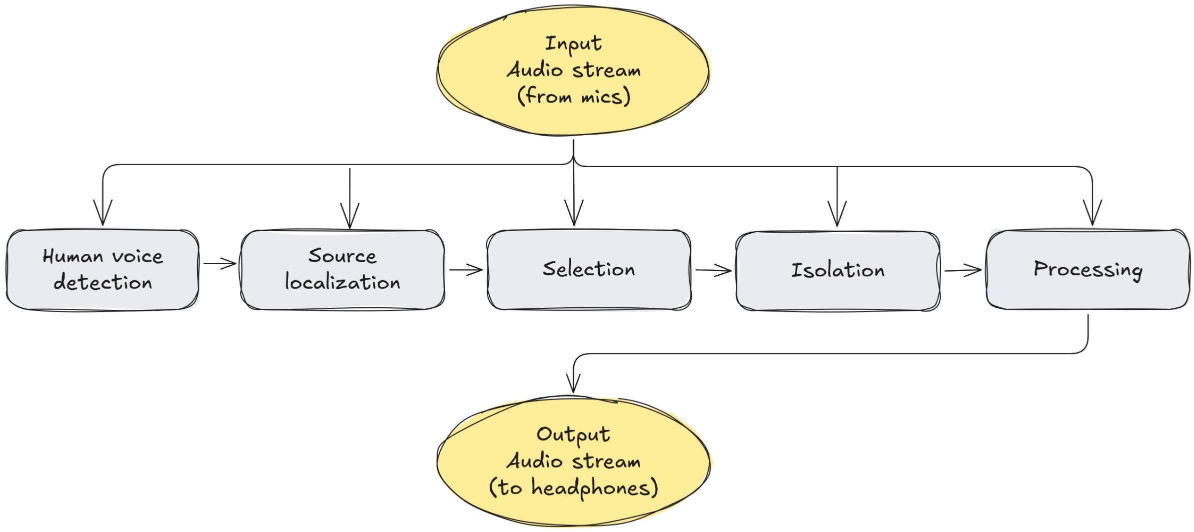

Figure 3: Software modules used for the processing of the audio stream

a dataset will be created as described in Section 4 which is essential for training or configuring the parameters of each of the modules described here.

In particular, the main software components are the following:

- **Human Voice Detection**: identifies the presence and the number of speakers. It is useful to start the following steps only when required (also for energy saving) and to enable appropriate algorithms depending on the number of audio sources.
- **Source localization**: estimates and tracks over time the positions of the speakers in the 3D space.
- **Selection**: this module allows a (semi)-automatic selection of which speaker the interest of the subsequent modules should fall on.
- **Isolation**: outputs an audio stream where only the voice of the selected speaker is audible.
- **Processing**: depending on the application, the final step could be a speech-to-text translation, a re-equalization, a frequency shift or any useful task.

3.3. Use Cases

The proposed use of AURALYS focuses on two application domains: a general workplace setting and a specific learning environment. Both contexts offer the possibility to detect sound features and to develop an appropriate response model, enabling the anticipation of suitable scenarios.

The first context is the general workplace, where an individual operates within a defined environment with known characteristics. In such settings, relevant auditory stimuli that are difficult to discriminate may occur sporadically or originate from directions that are challenging to localize.

The second context is educational, where multiple individuals interact simultaneously, and activities vary in nature. In this scenario, it may be particularly useful to isolate one voice from others, for example, by filtering based on intensity, timbre, or direction of origin.

In the following, we further discuss some real-world use-cases.

Difficulties. Several everyday environments present continuous and uniform background noise, intermittently punctuated by other sounds, only some of which are relevant. The relevance of these sounds may depend on their type (*e.g.*, equipment alarms) or their direction (*e.g.*, voices in crowded environments, such as communication among workers, or auditory signals perceived by drivers). In educational settings, such as classrooms, group activities often involve multiple overlapping voices and background noise, making it difficult for students to understand the teacher, or conversely, to follow what peers are saying when the teacher is interacting with others. This issue also extends to informal moments, such as recess or time spent in the schoolyard, where the ability to selectively focus on a specific auditory stimulus becomes challenging. In university contexts, similar difficulties can arise during collaborative activities, such as group discussions, or study sessions, where multiple simultaneous conversations may interfere with effective communication and participation.

Disorders. Here, the use case focus is on the case of ADHD, in which individuals are more prone to distraction during tasks due to difficulty in inhibiting irrelevant auditory stimuli – such as background voices or environmental sounds – and in shifting attention efficiently between stimuli. To support attentional performance in individuals with ADHD, several strategies can be adopted. One approach involves filtering or reducing the volume of non-relevant sounds or voices, based on characteristics such as direction, intensity, or timbre. This can help the individual maintain focus on the primary auditory stimulus, such as the teacher's voice or the voices of peers seated nearby or directly in front. Another strategy is to enhance or emphasize target voices within noisy environments – for example, during group discussions held outdoors – so that relevant speech stands out from background noise. A further option is to suppress or minimize all environmental sounds to create an artificially quieter setting. This can support concentration on cognitively demanding tasks, such as reading or studying, in otherwise noisy environments.

Disabilities. The performance of existing hearing aids can be improved by integrating their current software with the specific capabilities offered by the proposed software, without requiring the use of AURALYS glasses. Additionally, an even greater enhancement can be achieved by combining the software with AURALYS glasses, which provide directional sound capture from the surrounding environment, thereby further refining the device's ability to process relevant auditory stimuli.

4. The Verse framework

The "Virtual Environment for Rendering of Speech Emissions" (VERSE) framework[6] contains a platform to generate synthetic datasets of voice recordings and real environment characterization measurements. Among the others, the main goal is to study the advantages of an array of microphones versus binaural audio signals, in the context of real embedded devices and including machine learning algorithms for signal processing, with a particular focus on human voices.

VERSE is based on the abstraction of main components for an audio scene: voice sound sources from speakers, one listener and reverberation generated by the environment itself, meaning the room hosting speakers and listener. Specific to the definition of a scene is the concept of motion: the scene defines how sound sources are placed around the listener and how they move in space.

The basic resources defined and used in the VERSE framework are defined as follows.

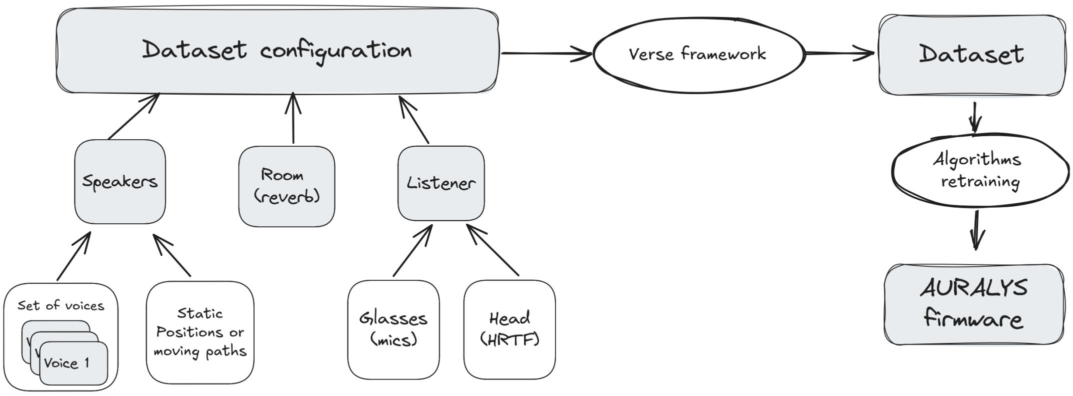

Figure 4: The VERSE framework for application-specific dataset generation

[6] https://github.com/iot-unimore/verse

- **Speakers**: these are audio recordings in a digital format capturing a single speaker recorded on a single track (mono) in a non-reverberant room (or as much low reverberation as possible). The absence of reverberation is important since the scene itself will define the type of room reverberation that must be applied. Each source can be static (does not move in space during playback) or dynamic (will move along a specific path in space), defining an audio scene.
- **Room**: environment defines early and late reflections of sounds arriving from walls and objects. The sum of all sound reflections to the listener's head (and microphones) defines the final audio perceived by the subject. Multiple techniques have been developed to properly measure the impulse response of a room, using energetic "impulsive" stimulus and sine sweep [24].
- **Listener**: the current version provides for a single listener. The listening subject is assumed to be at the centre of the three-dimensional reference system adopted to model the source paths. A listener can be defined as the coupling between a head and a pair of AURALYS glasses. The most important item contained in the listener definition is the HRTF of the microphone array (binaural or multi-microphone or both). The HRTF function is stored using the SOFA format (Spatially Oriented Format for Acoustics) [25] defined by the Audio Engineering Society (AES) in a specific standard (AES69-2022: AES standard for file exchange - Spatial acoustic data file format) [26].
- **Dataset configuration**: the dataset configuration contains all the information for the audio scene definition, *i.e.*the mix of speakers, room and listener.

Given a dataset configuration, the VERSE framework allows to render a complete dataset of dynamic scenes, moving the sources along a pre-defined path, thanks to the convolution engine by 3D TuneIn.

This flexibility of composing virtual audio scenes by swapping the resources into a definition is the essence of the VERSE dataset and framework: it is possible to generate a wide set of data with precise, repeatable positioning and with reference files (ground truth), skipping the cost and reducing the time for laborious recording sessions from real-life audio setup.

VERSE is also modular: the resource definition abstraction is done at a file level using YAML files to describe each resource with a common syntax and folder structure. This allows to add resources (sounds, heads or rooms) from other dataset to expand the possibilities for final audio generation.

5. Conclusion

In conclusion, the ability to perceive and selectively attend to relevant sounds in complex acoustic environments is a critical challenge, especially for individuals facing age-related, sensory, or neurodevelopmental vulnerabilities. These challenges are increasingly evident in everyday contexts such as schools and workplaces, where noisy conditions can significantly impair communication and learning. In these scenarios, the proposed AURALYS smart glasses represent a promising technological solution to address these issues. By enabling real-time directional sound capture and selective auditory filtering, AURALYS has the potential to enhance auditory perception and attention in noisy settings. The complementary VERSE framework supports this effort by providing targeted datasets optimized for the device's embedded processing constraints, facilitating efficient and context-specific sound source localization.

Future work will focus on refining hardware design, improving model generalization, and conducting extensive user-centered evaluations to fully realize the system's potential in promoting inclusive, accessible environments that support diverse auditory needs. This interdisciplinary approach, grounded in pedagogical principles and technological innovation, aims to improve quality of life for all individuals, particularly those who experience auditory vulnerabilities.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] R. Vakili, S. Vakili, M. Ajilian Abbasi, S. Masoudi, Overcrowded classrooms: Challenges, consequences, and collaborative solutions for educators: A literature review, Medical Education Bulletin 5 (2024) 961–972. doi:10.22034/meb.2024.492269.1103.

[2] D. Savegnago, L. Franz, M. Gubernale, C. Gallo, C. de Filippis, G. Marioni, E. Genovese, Learning disabilities in children with hearing loss: A systematic review, American Journal of Otolaryngology 45 (2024) 104439. doi:https://doi.org/10.1016/j.amjoto.2024.104439.

[3] R. Carlucci, E. Martinelli, P. Sapone, A. Cotroneo, The sound of silence: quanto il cervello non sente, ACSA MAGAZINE (2024). URL: https://www.acsamedical.it/the-sound-of-silence-quando-il-cervello-non-sente/.

[4] G. Bérard, Hearing Equals Behavior, New Cannan, Conn. : Keats Pub, 1993.

[5] A. Diamond, Executive functions, Annual Review of Psychology 64 (2013) 135–168. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-psych-113011-143750. doi:https://doi.org/10.1146/annurev-psych-113011-143750.

[6] B. Conte, G. M. Marzocchi, Specific executive function profile of children with adhd, learning disabilities or odd; [profili specifici di funzioni esecutive nei ragazzi con adhd, dsa o dop], Psicologia Clinica dello Sviluppo 24 (2020) 401 – 436. doi:10.1449/98293, cited by: 1.

[7] A. E. Marimpietri, M. C. Carmignani, A. Graziani, E. Sechi, Profili neuropsicologici e funzioni esecutive nei bambini con disturbo da deficit di attenzione/iperattività (adhd) e disturbo specifico di apprendimento (dsa) 79 (2012) 159–177.

[8] E. Cano, H. Lukashevich, Selective hearing: A machine listening perspective, in: 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP), 2019, pp. 1–6. doi:10.1109/MMSP.2019.8901720.

[9] T. Bent, I can't hear you without my glasses, J. Acoust. Soc. Am. 157 (2025) R5–R6.

[10] J. F. Culling, E. F. C. D'Olne, B. D. Davies, N. Powell, P. A. Naylor, Practical utility of a head-mounted gaze-directed beamforming system, The Journal of the Acoustical Society of America 154 (2023) 3760–3768. doi:10.1121/10.0023961.

[11] T. Fischer, M. Caversaccio, W. Wimmer, Multichannel acoustic source and image dataset for the cocktail party effect in hearing aid and implant users, Scientific Data 7 (2020) 440. URL: https://doi.org/10.1038/s41597-020-00777-8. doi:10.1038/s41597-020-00777-8.

[12] C. Gaultier, T. Goehring, Recovering speech intelligibility with deep learning and multiple microphones in noisy-reverberant situations for people using cochlear implants, The Journal of the Acoustical Society of America 155 (2024) 3833–3847. doi:10.1121/10.0026218.

[13] J. Barker, E. Vincent, N. Ma, H. Christensen, P. Green, The pascal chime speech separation and recognition challenge, Computer Speech Language 27 (2013) 621–633. doi:https://doi.org/10.1016/j.csl.2012.10.004, special Issue on Speech Separation and Recognition in Multisource Environments.

[14] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, M. D. Plumbley, Detection and classification of acoustic scenes and events, IEEE Transactions on Multimedia 17 (2015) 1733–1746. doi:10.1109/TMM.2015.2428998.

[15] H. Steffens, S. van de Par, S. D. Ewert, The role of early and late reflections on perception of source orientation, The Journal of the Acoustical Society of America 149 (2021) 2255–2269. doi:10.1121/10.0003823.

[16] E.-G. N. Erwin Meyer, Physical and Applied Acoustics: An Introduction, Academic Press, New York, 1972.

[17] V. Pulkki, J. HUOPANIEMI, Analyzingvirtual sound source attributes using a binaural auditory model, Journal of the Audio Engineering Society. Audio Engineering Society 47 (1999).

[18] M. Zhang, W. Zhang, R. Kennedy, T. Abhayapala, Hrtf measurement on kemar manikin, Annual Conference of the Australian Acoustical Society 2009 - Acoustics 2009: Research to Consulting (2009) 10–17.

[19] M. Rothbucher, K. Veprek, P. Paukner, T. Habigt, K. Diepold, Comparison of head-related impulse

response measurement approaches, The Journal of the Acoustical Society of America 134 (2013) EL223–EL229. doi:10.1121/1.4813592.

[20] A. Farina, Simultaneous measurement of impulse response and distortion with a swept-sine technique (2000).

[21] Hrtf measurements with recorded reference signal, in: 129th Audio Engineering Society Convention 2010, 129th Audio Engineering Society Convention 2010, 2010, pp. 533–540. 129th Audio Engineering Society Convention 2010 ; Conference date: 04-11-2010 Through 07-11-2010.

[22] M. Cuevas-Rodriguez, D. Gonzalez-Toledo, A. Reyes-Lecuona, L. Picinali, Impact of non-individualised head related transfer functions on speech-in-noise performances within a synthesised virtual environment, The Journal of the Acoustical Society of America 149 (2021) 2573–2586. doi:10.1121/10.0004220.

[23] D. O'Connor, J. Kennedy, An evaluation of 3d printing for the manufacture of a binaural recording device, Applied Acoustics 171 (2021) 107610. URL: http://dx.doi.org/10.1016/j.apacoust.2020.107610. doi:10.1016/j.apacoust.2020.107610.

[24] R. San Martín, M. Arana, J. Machín, A. Arregui, Impulse source versus dodecahedral loudspeaker for measuring parameters derived from the impulse response in room acoustics, The Journal of the Acoustical Society of America 134 (2013) 275–284. URL: http://dx.doi.org/10.1121/1.4808332. doi:10.1121/1.4808332.

[25] A. E. S. (AES), Sofa (spatially oriented format for acoustics), 2022. URL: https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics).

[26] A. E. S. (AES), Aes69-2022: Aes standard for file exchange - spatial acoustic data file format, 2022. URL: https://www.aes.org/publications/standards/search.cfm?docID=99.