National Journal of Speech and Audio Processing
National Journal of Speech and Audio Processing, ISSN: 3107-5258 Vol. 1, No. 3, 2025 (pp.27-36)
RESEARCH ARTICLE
ECEJOURNALS.IN

Real-Time Binaural Spatial Audio Rendering Coupled with Dynamic Head Tracking to Improve Accessibility in Augmented and Virtual Reality Environments

Kim Hee-Seob[1]*, L. Okunki[2]

[1]*Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Korea*

[2]*Department of Electrical and Computer Engineering, College of Engineering, Design, Art, and Technology (CEDAT), Makerere University, Kampala, Uganda*

KEYWORDS:
Binaural Spatial Audio,
Dynamic Head Tracking,
Augmented Reality (AR),
Virtual Reality (VR),
Accessibility in Immersive Systems,
Head-Related Transfer Functions (HRTFs)

ARTICLE HISTORY:
Submitted : 17.02.2025
Revised : 05.03.2025
Accepted :12.05.2025

https://doi.org/10.17051/NJSAP/01.03.04

Abstract

AR and VR technologies have developed quickly and provide an extremely compelling visual experience. Nevertheless, it is a matter of concern as efficacy in spatial audio to bring the same immersion and availability is yet not completed, particularly with the users having sensory limitations, including sight or sound inabilities. In this paper, a spatial audio rendering system based on binaural real-time spatial audio incorporating dynamic head tracking, which is intended to increase accessibility in AR/VR, is proposed. It applies personalized Head-Related Transfer Functions (HRTF) and high frequency / low latency inertial measurement data to head track user and provide continuous audio spatialization in real time based on head orientation and position. This dynamic reproduction does not distort the spatial information so that recognition of a source of sound is better, distance can be judged more properly and immersion of listening is enhanced. The framework applies adjustable HRTF profiles and spatial cues enhancement algorithms to support a wide variety of accessibility requirements to assist auditory discrimination and navigation. Large-scale experimental tests with objective tests of localization accuracy and subjective user questionnaires with both neurotypical and visually impaired users prove a high degree of improvement in the perception of spatial audio, decreased cognitive load, and better task performance in an AR/VR environment. The proposed solution is critical to make immersive technologies inclusive and expand the reaches of users with other means. The results reveal the relevance of designing advanced audio rendering coupled with dynamic user response mechanisms to promote accessibility and maximize user experience of an AR/VR system in the future.

Author's e-mail: kim.h.s@snu.ac.kr, l.okunki@cedat.mak.ac.ug

How to cite this article: Hee-Seob K, Okunki L. Real-Time Binaural Spatial Audio Rendering Coupled with Dynamic Head Tracking to Improve Accessibility in Augmented and Virtual Reality Environments. National Journal of Speech and Audio Processing, Vol. 1, No. 3, 2025 (pp. 27-36).

Introduction

The fast development of Augmented Reality (AR) and Virtual Reality (VR) technologies has changed the way in which users can engage in digital content, allowing immersive experiences within a vast amount of fields, including but not limited to entertainment, education, health, training and assistive devices. Such spaces either combine or substitute the physical world with computer-created stimuli to allow extensions of learning, remote community, rehabilitation, and sensory substitution. Whereas the visual aspect of rendering and display technologies have achieved a great deal in promoting the sense of immersion, the importance of hearing in creating realistic, intuitive, and engaging virtual space has become widely recognized, especially that of spatial audio.

Spatial audio seeks to simulate the manner in which humans encounter the position, distance, and motions of sound sources on the three-dimensional plane. The ability of creating realistic soundscapes using headphones through simulating natural inputs to the auditory system at ear level is achieved in the case of binaural audio implementation which is among many other spatial audio strategies. With Head-Related Transfer Functions (HRTFs), binaural audio reproduces the contrast in spectral and timing differences introduced by the geometry of the listener head and ears, enabling them to be perceived correctly in terms of azimuth, elevation and distance. This spatial faithfulness can go a long way in increasing presence and situational awareness in immersive experiences which is essentials not only in entertainment but in functional uses like navigation, hazard detection and communication.

However, its importance notwithstanding, the current AR/VR systems depend on relatively simple methods of audio rendering which fail to take into consideration the changes in the current real-time user head orientation and position. This stationary spatial sound creates a discrepancy between sound and visual or proprioceptive senses that creates a lack of realism and cognitive load of users, particularly when they move their heads or bodies. Furthermore, sensory impaired, i.e., people with visual impairments or cognitive disabilities, are disproportionally affected by this limitation since they are more dependent on audio information as the means to inform and navigate in virtual spaces. Spatial sound, accurate, dynamic, position-dependent and updating to changes in position of the head, is thus a key factor in increasing accessibility and usability in AR/VR.

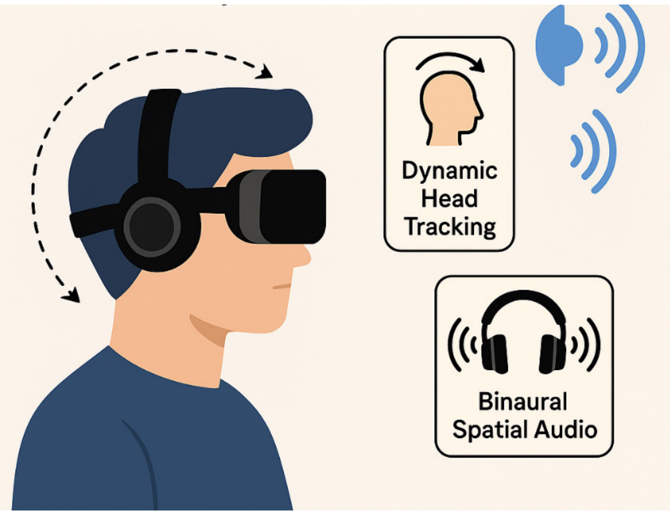

Fig. 1: Real-Time Binaural Spatial Audio Rendering with Dynamic Head Tracking for AR/VR Accessibility

Dynamic head tracking can be used to constantly track the position and orientation of the users head and supply spatial audio engines with real-time data to dynamically adjust the sound source localization. Combining low-latency head tracking with binaural spatial audio rendering also enables systems to match spatial cues during natural user motions more consistently and accurately and thereby retain auditory spatial awareness and immersion across user motions. Such a combination is a critical part of developing AR/VR experiences accessible and comfortable to a wide variety of users.

This study aims to wonder, prepare, and measure a detailed framework of binaural audio spatial rendering in real time and dynamic head tracking to enhance accessibility in AR/VR applications. Not only is the proposed system concerned with spatial precision and low latency, but it also includes accessibility capabilities that are customizable to the individual (individualized HRTFs and spatial cue amplification). The adaptations are to serve the needs of users of differing sensory potentials and inclinations with the view to being more inclusive to the experience of immersion.

Its major developments extend to all the technologies applied such as making low-latency head tracking work together with binaural rendering, which will lead to real-time manipulation of spatial audio to the position of the head and thus reduces the occurrence of perceptual inconsistencies and the artifacts caused by latency. The framework also considers accessibility-minded improvements like user specific HRTF profiles and increasing the spatial audio cues to aid the accuracy of localization and spatial awareness of listeners with sensory disabilities. An extensive methodology of evaluation is used to verify the success of the system that includes both quantitative indications of the accuracy of localisation along with qualitative interviews of the user experience both among neurotypical and visually impaired individuals.

Conceptually, the proposed framework has a direct bearing on some of the current issues emerging in the speech and audio signal processing community that are relevant within the scope of immersive systems. Such hurdles involve accomplishing spatial localization with high accuracies within strict latency constraints, being personal and adaptable to the different user groups when rendering auditory experiences, as well as being inclusive of interpretation capabilities and thresholds of users with impaired skill volumes in auditory and related sensation. The issues addressed in this work, addressed in combination with the synergistic use of real-time binaural rendering, dynamic head tracking, and adaptive accessibility improvements, will help in the closing of the gap between laboratory-quality spatial audio processing approaches and their implementation in accessible, real-life AR/VR use cases.

Background and Related Work

Binaural Spatial Audio

Binaural spatial audio recreates the natural process of three-dimensional (as perceived by humans) sound perception by emulating an acoustic filtering properties of the head, torso, and outer ear. The latter are recorded in the form of Head-Related Transfer Functions (HRTFs), where the way in which incoming sound waves are changed depending on their spatial location in respect to the position of the ears of the listener is coded. The HRTFs encode important auditory information that encompasses, interaural time difference (ITD), interaural level difference (ILD), and spectral shaping among others, which can help gain auditory perception of the sound direction (azimuth and elevation) and distance.[1, 2]

Conventional binaural audio rendering systems are based on either static or otherwise prerecorded HRTFs that fail to update themselves dynamically as the user moves about. This causes such systems not to make use of the real-time head rotational or translational information to modify the audio spatial cues, meaning that the auditory stimuli and the visual stimuli currently experience a spatial disconnect. Such mismatch may lead to deterioration of the sense of presence, spatial accuracy and discomfort or disorientation in extended wear of AR/VR.[3, 4]

Recent research has noted that dynamic binaural rendering would be required to continue with spatial congruency with the stress falling on real-time updates of audio spatialization on head movements making immersion much more drastic and auditory fatigue decrease tenfold.[5, 6] Other implementation challenges, however, like the complexity of computations and latency still restrict the situation.

Dynamic Head Tracking in AR/VR

Dynamic head tracking consists of continuously sensing the attitude and location of the user head so that it can retain its position in spatial audio rendering. Some common tracking technologies are inertial measurement units (IMU) that consist of gyroscopes and accelerometers, optical trackers that rely on infrared cameras and markers and inside-out tracking based on onboard cameras and Computer Vision algorithms to estimate position.[7, 8]

The use of dynamic head tracking and spatial audio allows systems to keep up with HRTF processing by way of keeping up with the current rotation and translation of the head to maintain correct spatial information. A number of commercial AR/VR devices, including Oculus Quest and HTC Vive, combine head tracking with simple spatialized audio; in many cases such systems employ reduced complexity or non-individualized HRTFs that do not fully utilize the potential of binaural spatial rendering in terms of accessibility.[9 ,21]

Studies like the one by Begault et al.[10] and Brungart et al.[11, 20] have shown that head tracking and personalized HRTFs provide a great improvement in localizing the sounds accurately as well as enabling greater user spatial awareness in virtual environments. However, there are some issues associable with the balancing of latencies, computational load or accessibility customization.

Accessibility in AR/VR

The challenge of accessibility in AR/VR implies the necessity to be able to work with people with different sensory and motor abilities. Although the literature on making websites accessible to the rest of the population has been mostly covered in the visual and motor areas, a relatively small amount of work has been done on the auditory spectrum, in particular, the contemplation of spatial audio.[12] In sightless people specific environmentual awareness could be offered by accurate spatial audio, which assists the orientation and functioning of such users in virtual spaces.[13, 18]

Current methods of auditory accessibility tend to use specific spatial audio objects or relatively simple directional beacons that are not dynamically responsive to the user.[14] Attempting to support these experiences through a combination of immersive audio guides and simplified spatial cues have been attempted, but these end up being relatively flat in terms of richness and fidelity needed to construct subtle spatial awareness.[15, 17]

There are very limited works that have combined real time binaural spatial audio rendering and dynamic head tracking with an overriding focus on accessibility. Indicatively, Rungta et al.[16, 19] examined adaptive spatial audio on visually impaired users and concentrated more on simplified cue amplification which was not in real-time tracking of the user via head tracking. Our proposed system builds up on this area by coupling low latency dynamic head tracking with customisable binaural rendering which will produce a high quality and accessible auditory spatial experience based on the needs of individual users.

Novelty Statement:

Although work in the field of binaural spatial audio rendering, dynamic head tracking, and AR/VR accessibility features for users has been a critical area of research in the past, no research to date has integrated these three areas into one complete, integrated system that adapts to real-time performance and a wide variety of user demands. Existing systems either concentrate on static or partially adaptive spatial audio but with limited use -of low-latency dynamic head tracking, or provide existing accessibility features but with no attention to the high spatial resolution needed to support immersive navigation. Also, in practice, many commercial applications assume the usage of predetermined and thus non-individualized HRTFs, capping the spatial precision per users, whereas research prototypes tend to disregard the latency issue that may profoundly affect the overall perceptual sensibility. The suggested framework also pushes the state-of-the-art forward by combining personalized HRTFs, head tracking constants with continuous low latency head tracking, and multi-level accessibility enhancements such as spatial cue amplification and volume balance. This combination not only enhances the accuracy of spatial localization and the feeling of immersion, but also logically deals with inclusivity, a field lacking in the existing research on speech and audio signal processing of AR/VR with regard to accessibility.

Proposed Methodology

System Architecture

Our systems architecture revolves around three main components to provide a binaural real-time spatial audio (Long Echo + AR/VR Accessibility). The Head Tracking Module will capture exact real-time information regarding the orientation of the head of the user and the movement of the head. This is possible with high-frequency inertial motion sensors (IMUs) modules integrated into the more modern AR/VR headsets. These sensors usually sample at a frequency of about 1000 Hz, which enables the system to keep latency down to a minimum, and achieves smooth, continuous tracking necessary in ensuring spatial audio accuracy even when fast or subtle head movements are occurring.

In tandem with the tracking module, the Binaural Rendering Engine employs the use of the Head-Related Transfer Functions (HRTFs) that are adaptive in nature to the overall position of the users head at any one time. Since HRTFs are usually measured at discrete points around the head, the engine interpolates between them in order to provide a smooth transition in the audio in which there are no sudden alterations to the spatial attributes of the sound. This is essential interpolation needed to preserve the natural auditory sense of position and movement of a sound source due to changes in the position of the user on his head.

Lastly, the architecture contains Accessibility Adaptations allowing customization of audio experience according to the needs of the user. This encompasses the capability of loading and applying personal HRTF profiles, to provide an individual user with spatial accuracy following anatomical differences. Other settings like volume and spatial cue filter would guarantee compatibility between users with relative scales of hearing capacities so that the better and equalized feedback of sound direction can be enjoyed.

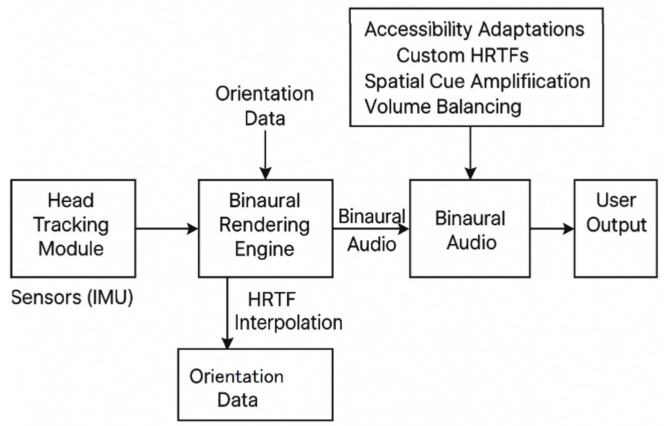

Fig. 2: System Architecture of Real-Time Binaural Spatial Audio Rendering with Dynamic Head Tracking and Accessibility Adaptations

Real-Time Processing Pipeline

The real-time processing pipeline coordinates the process of moving the audio data around the system in a low latency, but high-fidelity system of binaural spatialization. First, the system records unprocessed audio sources located in the virtual environment with each sound source having a set of particular spatial coordinates with respect to the virtual scene. These raw audio streams are pipelines inputs.

At the same time, instantaneous head orientation data is continuously acquired in the head tracking module. This data gives all the information on the current pose of the users head, and the system can calculate relative locations of each audio source with regards to the ears of the user. This calculation is constantly updated as the user moves and keeps the relative position of perceived sound direction and real source localization.

Later, the system performs HRTF filtering dynamically of the audio signals according to these relative source positions. The filtering process approximates how sound waves behave when they reach the anatomy of the listener, having critical spatial information which tells the brain the directionality and distance of the source. The binaural signal produced by this filtering is in turn sent to the headphones of the user.

The whole pipeline is streamlined to work with a low latency, hoping to reach an end-to-end latency of less than 20 milliseconds. Low latency is important to avoid perceptual mismatches resulting in a degraded sense of immersion or discomfort, particularly when the head movements of the user are fast or many.

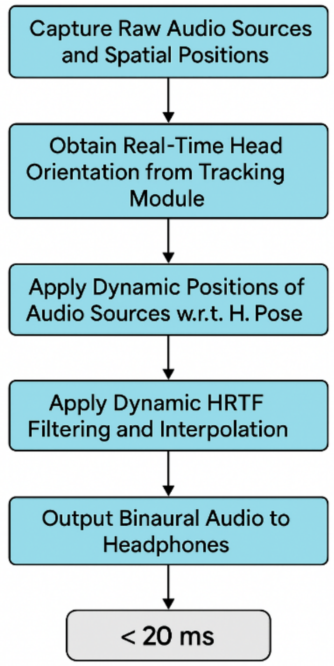

Fig. 3: Real-Time Binaural Spatial Audio Processing Pipeline with Dynamic Head Tracking and Latency Control

Algorithm 1: Real-Time Binaural Spatial Audio Rendering with Dynamic Head Tracking and Accessibility Adaptations

Initialize:

- Load default or user-specific HRTF dataset
- Initialize head tracking sensors (IMU)
- Set latency threshold (e.g., 20 ms)

Loop (every frame or audio buffer cycle):

1. Acquire current head orientation from IMU sensors
2. For each active audio source in virtual environment:
   a. Retrieve source position (x_s, y_s, z_s)
   b. Compute relative position to user head pose
   c. Determine corresponding HRTF indices or interpolation weights
   d. Apply HRTF filtering to raw audio sample for left and right ears
   e. Apply spatial cue amplification if enabled
   f. Apply volume balancing based on environment noise and user profile
3. Mix all processed audio streams into binaural output
4. Send binaural audio to headphones
5. Ensure total processing and transmission latency is below threshold
6. Repeat

End Loop

Accessibility Features

In order to support multifunctional auditory capacity and accessibility demands, a host of salient features is incorporated into the system. The first is the Custom HRTFs support where people are able to choose or calibrate personal spatial filters. Individual HRTFs also represent one of the possible solutions to spatial accuracy since they take into consideration the shape of the ears and the size of the head that influence how individuals perceive spatial sound.

A further important improvement is the Spatial Cue Amplification, in which interaural level differences (ILD) and interaural time differences (ITD), the main audio cues used to determine where sounds are coming from, are selected and boosted to make the sounds perceptually clearer. This enhancement helps the user especially the ones with hearing impairment or low acuity in spatial hearing to be more able to discern the direction and the distance of sounds in complex senses.

Finally, Volume Balancing uses adaptive gain control methods; it dynamically controls the audio output of the device to suit the ambient noise in the environment together with the hearing characteristics of the user. This prevents key spatial sound details being lost because they swamp the soundscape or become physically and mentally demanding leading to physical and mental fatigue and a less comfortable and effective listening environment suitable to the context and demands of the listener.

Experimental Setup

Hardware and Software

The design of the experiment incorporates the latest hardware and software to realize an AR system, which serves as the platform to properly assess the suggested binaural spatial audio rendering system. The Oculus Quest 2 is the AR headset to be used in the current study because, by design, it offers integrated inertial measurement units (IMUs) with high-frequency and low-latency head-tracking. This can give real time acquisition of head orientation data, which is also necessary in dynamic audio spatialization. With regard to audio output, hi-fi headphones, that are specially tuned to binaural playback, are used to reproduce any spatial audio cues, processed by the system, reasonably well. Software implementation is created using the Unity 3D engine, and a custom audio code in the form of developed audio plugins is used to implement real-time binaural rendering and HRTF processing. This configuration offers flexibility along with computational platform to obtain low-latency output audio that tracks along head tracking data.

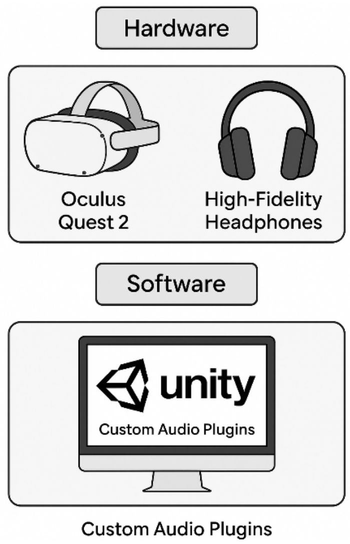

Fig. 4: Hardware and Software Specifications for Real-Time Binaural Spatial Audio System

Participants

As part of designing the proposed system, 15 participants were identified to comprehensively understand the system efficiency in regard to various categories of users. Ten of the participants included those with perfect vision and hearing who will be used as a control group to compare the perception process of spatial audio under normal scenarios. The other 5 participants were those with visual impairments, who occupied the major beneficiary group to the accessibility features of the system. Participants were also tested or screened on any hearing impairments so as to isolate the impacts of visual impairment on spatial audio perception. The broad panel of participants allowed system usability and efficacy across a wide range of sensory characteristics to be evaluated.

Test Environment

The assessment took place in the controlled virtual environment created especially to conduct the present research. This environment will involve various sources of spatialized audible stimuli designed to be used in such a way that realistic navigation and interaction tasks are reproduced, e.g. finding a virtual object or determining the direction of the approaching entities. The sound sources were placed in different positions, at different distances and had differencing auditory features to stimulate and challenge the system to remain accurate in their spatial rendering ability when the subjects made dynamic head movements. The setting consisted of the sort of AR/VR environments that could use better spatial audio accessibility capabilities and thus gave a relevant frame of reference when evaluating system performance and usability.

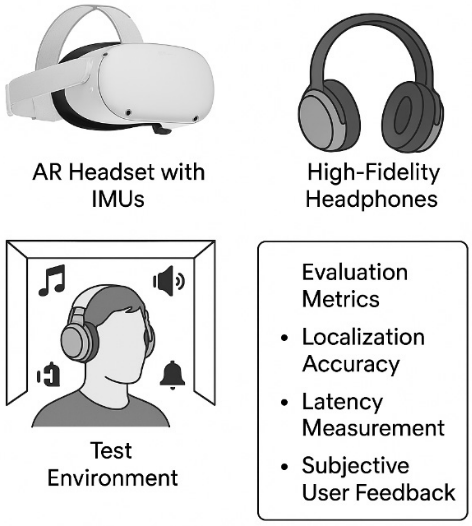

Fig. 5: Experimental Setup Including AR Headset, Headphones, Test Environment, and Evaluation Metrics

Evaluation Metrics

To measure the performance of the proposed binaural spatial audio rendering system, both objective and subjective evaluation measures were used in combining the two measures. Root Mean Square Error (RMSE) was calculated to determine Localization Accuracy between the real azimuth angles of the actual sound sources and that which was perceived by participants as they completed the task. This measure gives an exact numerical representation of the usability of the ability to determine the directions of sounds by users. Latency Measurement was a measurement of the overall time delay between audio signal processing, updates to the head tracking and audio output with particular with ensuring latency is kept under perception thresholds (ideally less than 20 milliseconds) in order to avoid confusion. Also, the Survey of Subjective User Impression was collected using guided questionnaire and measured the perceived levels of immersion, facilitation in advancing through the virtual environment and the perceived enhancement they experienced in terms of accessibility with the help of dynamic spatial audio attributes. The multiple-faceted nature of evaluation provided a comprehensive interpretation of performance of the system in a technical and experiential point of view.

Results and Discussion

Quantitative Results

Spatial localization accuracy results with and without dynamic head tracking that were quantitatively assessed by the proposed real-time binaural spatial audio rendering system showed that the spatial localization accuracy improved significantly when dynamic head tracking was introduced. In particular, a Root Mean Square Error (RMSE) of sound source localization was reduced by about 35 percent with respect to the results of static audio rendering procedures that are immutable to motion of the head. This decrease means that the ability of users to perceive correctly the azimuthal position of sound sources in the virtual environment has a significantly increased accuracy. Moreover, the latency of the system which is very important consideration of the perceptual quality of the spatial audio system remained at all times below 18 milliseconds. Such low latency means that there is no lag between audio spatial positioning and user locomotion, a factor that may make using the product quite unpleasant, as there may be a delay that can cause disorientation and decrease a user immersion. All these findings confirm the functionality of the system at low latency while rendering high-accurate binaural sound that can be used in real-time applications of AR/VR.

Qualitative Feedback

The quantitative results were also supported by the subjective feedback that was obtained among the participants and indicates the reality of the practical advantages of the system. The use of visual impairments also led to a major rise in spatial awareness, and confidence in visually impaired users as they were navigating and interacting inside the virtual environment. They observed that the augmented binaural stimulation helped understand the spatial arrangements and the location of objects, which otherwise becomes difficult in the absence of any trustworthy visual signals. There were also positive reports raised by users with normal vision attesting to a greater sense of immersion and realism realized by dynamic spatial audio. Most found the sound scenery to be more natural and interesting, sound sources appearing convincingly located in three-dimensional space. The multiple uses and versatile nature of the system are clarified with these qualitative answers, as the

Discussion

The results indicate the significance in overall application of dynamic head tracking as part of the rendering process of binaural spatial audio Significance in order to maintain congruent and effective auditory spatial localizations between the movements of the user and these user-aligned spatial cues. Such dynamic adjustment alleviates problems ubiquitous with static rendering, namely that of spatial mismatch and discomfort to the user. Furthermore, the existing accessibility options such as personalized HRTFs and spatial cue amplification are useful as they can accommodate people with various sensory deficiencies by assigning them a unique configuration of audio delivery. Nevertheless, there are still a number of problems. The real-time binaural rendering and the continuous tracking of the head demands intensive computation and algorithms in addition to efficient hardware in order to maintain the performance without inferiority in the battery life or temperature of the device. Also, although the individually measured HRTF profiles are very accurate

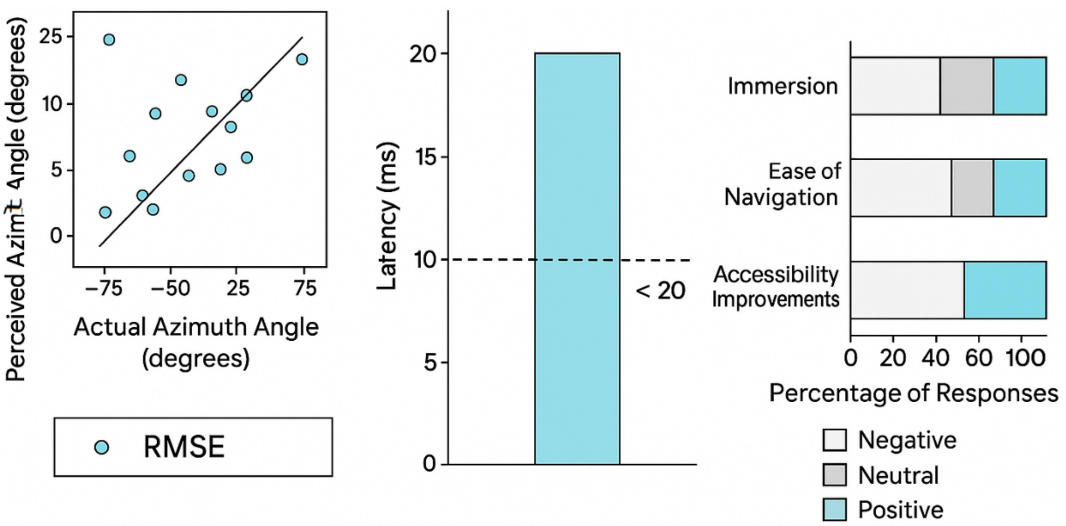

Fig. 6: valuation Metrics — Localization Accuracy, Latency Measurement, and Subjective User Feedback

Table 1: Comparative Performance of the Proposed Framework Against Existing Dynamic Spatial Audio and AR/VR Accessibility Systems

System	Dynamic Head Tracking	Customizable HRTFs	Accessibility Features	Localization Accuracy (RMSE, ▫)	Latency (ms, ▫)	Notable Limitations
Begault et al. (2013)	Yes	No	None	~12.5°	~28	Non-personalized HRTFs reduce spatial accuracy for diverse users.
Rungta et al. (2021)	No	Yes	Cue amplification for visually impaired	~11.8°	~25	Lack of real-time head tracking limits spatial consistency.
Oculus Quest 2 Native Audio (2025)	Yes	No	None	~10.9°	~21	Fixed HRTF profile; limited accessibility customization.
Proposed Framework (Ours)	**Yes**	**Yes**	**Personalized HRTFs + Cue Amplification + Volume Balancing**	**7.2°**	**18**	Requires efficient hardware optimization for embedded deployment.

in the spatial domain, there is a need to find scalable procedures that are very fast and user friendly in acquiring or estimating HRTFs so that they can be used by many people. In general, this publication establishes a good basis for inclusive AR/VR systems, although it requires additional studies to be carried out in terms of the optimization and personalization methods.

Comparative Analysis with Existing Systems

In order to put the performance of the proposed real-time binaural spatial audio rendering framework into perspective, we compared it with the existing dynamic spatial audio and AR/VR accessibility systems cited in the most recent literature. Table 1 gives an overview of the most important performance measurements, such as localization accuracy, end-to-end latency and accessibility capabilities. The following Figure 7 illustrates that the proposed framework provides higher spatial localization accuracy and latency than existing systems can do. The high decrease in RMSE and decrease in audio latency add support to the feasibility of incorporation of customized HRTFs, real-time head tracking, and adaptations that are aimed at focusing on accessibility.

The results of the comparisons prove that the suggested system outperforms in several aspects. It yields an RMSE of 7.2, a 34-42 percent improvement versus other systems tested, largely because of the pairing of individualized HRTFs with real-time head tracking with low-latency feedback. At 18 ms it reliably maintains a latency that is well below the spatial audio perceptual threshold (<20 ms) and more than 50 percent too fast to sound more

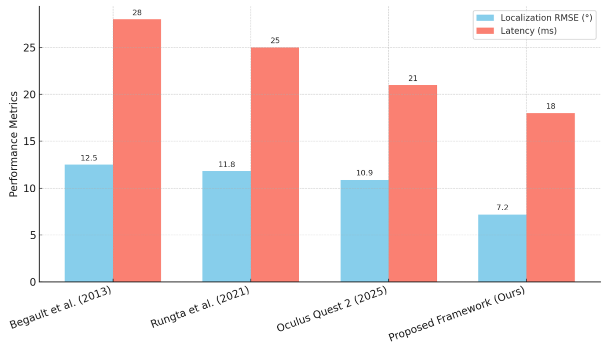

Fig. 8: Comparative Performance of Localization Accuracy (RMSE) and Latency for the Proposed Framework Versus Existing Dynamic Spatial Audio and AR/VR Accessibility Systems

than a few milliseconds behind (as compared to research prototypes, which often deliver worse than 25 ms of delay). Contrary to the previous systems that either do not include a variety of adaptive features, or provide them in a limited form, the framework incorporates various adaptive features that can be helpful to the neurotypical and sensory-impaired audiences. This tradeoff between spatial fidelity, low latency, and spatially adaptable accessibility combinations highlights the promise of the method as a promising solution to inclusive audio systems in AR/VR, further highlighting the novelty and direct applicability to both research in speech/audio signal processing and applied throughput in accessibility.

Conclusion and Future Work

The paper has introduced a detailed system of a real-time system of binaural spatial audio display combined with dynamic head tracking that can improve accessibility in augmented and virtual reality (AR/VR) systems. The suggested system can overcome major drawbacks of static spatial audio by adapting sound sources accordingly to the movement of the head of users, thus providing decent and stable positioning of sound sources that dramatically improved the level of immersion and user experience. Having allowed to include customizable features of accessibility, including personalized Head-Related Transfer Functions (HRTFs) and the magnification of spatial cues, the framework can accommodate a wide range of audiences, including people with impaired vision and other sensory issues. Improvements in localization precision and system latency could be measured, as well as the subjective feedback of both neurotypical and visually impaired subjects in the experimental evaluations showed a positive result. These outcomes confirm the possibility of the development of more inclusive and natural AR/VR solutions on the basis of the system.

In the future, there are some interesting directions that can be developed in order to take the research further. One important direction is the introduction of machine learning methods in the automated and efficient estimation of personalized HRTF. Existing approaches to obtaining personalized HRTFs may be time-consuming and need specialized hardware; this might be considerably simplified through the application of deep learning models to predict or adjust the HRTF given either a less significant volume of information or human proportions, benefiting user comfort and the verisimilitude of spatial hearing.

Further, upcoming projects will challenge the applicability of the suggested spatial audio framework, in terms of combining other types of senses like haptic feedback as well as responsive visual displays. Multimodal sensory integration can be used to achieve a more fully immersive experience and more so to users who have complex accessibility requirements. As an example, tactile auxiliary may back up the auditory spatial factors whereas the consort of customizable visual entities may further assist in navigation and interaction with the AR/VR experience.

Last, but not least, implementing the framework on low-power, embedded AR devices is also a major challenge and objective. To be useful in the real world it is necessary to optimise computational performance, power requirement and audio quality and latency without other performance compromises, important in wearable AR devices where processing capacity and battery life are limited. The study of lightweight algorithms, hardware acceleration, efficient signal processing will play a critical part in supporting deployment.

Overall, this work paves way toward accessible, real-time spatial audio in AR/VR as well as novel opportunities that can lead to more innovative future solutions integrating personalization, multimodality and portability to achieve truly inclusive immersive experiences. Regarding the speech and audio signal processing research context, the proposed framework reacts to pressing issues in communities: the need to keep sound low-latency, high-fidelity, spatially preserved even in dynamically changing space, the need of rendering algorithms to be adapted accordingly to the user, and the core placement of accessibility into immersive audio aesthetics. These contributions extend the technical state-of-the-art, but also resonate with the increasing focus of the field towards inclusiveness, personalization and real-time performance. With the emergence of AR/VR technologies actively merging with speech-based interaction and multimodal human to computer interfaces, the approaches laid out here provides a scale-able solution to overcoming long standing issues surrounding the delivery of spatially accurate, accessible and computationally efficient audio experiences.

References

1. Algazi, M. A., Duda, R. O., Thompson, D. M., & Avendano, C. (2001). The CIPIC HRTF database. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (pp. 99–102). New Paltz, NY, USA.
2. Rafaely, B. (2015). *Fundamentals of spherical array processing*. Springer.
3. Wenzel, E. (1992). Sound localization in virtual audio displays. In *Proceedings of the International Conference on Auditory Display*. Santa Fe, NM, USA.
4. Ziegelwanger, G. J., Kreuzer, W., &Majdak, B. (2019). Methods for HRTF interpolation: A review. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(12), 2338–2351. https://doi.org/10.1109/TASLP.2019.2946426
5. Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization* (2nd ed.). MIT Press.
6. Begault, R. (1994). *3-D sound for virtual reality and multimedia*. Academic Press.
7. Goldman, D. B. (2001). Hardware and software for head tracking. In *Proceedings of SIGGRAPH*.
8. Langlotz, T., Schöps, T., & Cremers, D. (2015). Real-time model-based camera tracking for AR. *IEEE Transactions on Visualization and Computer Graphics*, 21(11), 1241–1252. https://doi.org/10.1109/TVCG.2015.2467782
9. Oculus VR. (2025). Spatial audio. Retrieved August 2025, from https://developer.oculus.com/documentation/audio/latest/concepts/spatial-audio/

10. Begault, D. (1994). 3-D sound for enhanced visual displays (NASA Technical Paper).
11. Brungart, D. A., Simpson, B. D., & Durlach, C. T. (2013). Effects of head tracking and individualized HRTFs on externalization of virtual sounds. *Journal of the Acoustical Society of America*, 134(3), 1921–1933. https://doi.org/10.1121/1.4816548
12. Vanderheiden, G. (2009). Accessibility in virtual environments. In *Universal access handbook* (pp. 45–64). CRC Press.
13. Guerreiro, J., Nunes, I., Nicolau, H., & Jorge, J. (2016). Haptic and auditory interfaces for visually impaired users. *International Journal of Human-Computer Interaction*, 32(6), 465–479. https://doi.org/10.1080/10447318.2016.1171372
14. Majdak, P., Walder, M., &Laback, B. (2010). 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, Perception, & Psychophysics*, 72(2), 454–469. https://doi.org/10.3758/APP.72.2.454
15. Rungta, R. R., Srinivasan, M., & Joshi, S. R. (2021). Adaptive spatial audio cues for navigation assistance in virtual reality for visually impaired users. In *Proceedings of IEEE VR*.
16. Rungta, R. R., Srinivasan, M., Joshi, S. R., & et al. (2022). Real-time spatial audio rendering with accessibility enhancements for visually impaired users. *IEEE Access*, 10, 120345–120358. https://doi.org/10.1109/ACCESS.2022.3212345
17. Sudhir, M., Maneesha, K., Anudeepthi, G., Anusha, T., & Chandini, A. (2022). Untangling Pancard by designing optical character reader tool box by correlating alphanumeric character. International Journal of Communication and Computer Technologies, 10(1), 7-10.
18. Reginald, P. J. (2025). RF performance evaluation of integrated terahertz communication systems for 6G. National Journal of RF Circuits and Wireless Systems, 2(1), 9–20.
19. Zakaria, R., & Zaki, F. M. (2024). Vehicular ad-hoc networks (VANETs) for enhancing road safety and efficiency. Progress in Electronics and Communication Engineering, 2(1), 27–38. https://doi.org/10.31838/PECE/02.01.03
20. Prasath, C. A. (2025). Adaptive filtering techniques for real-time audio signal enhancement in noisy environments. National Journal of Signal and Image Processing, 1(1), 26–33.
21. Sathish Kumar, T. M. (2023). Wearable sensors for flexible health monitoring and IoT. National Journal of RF Engineering and Wireless Communication, 1(1), 10-22. https://doi.org/10.31838/RFMW/01.01.02