ISSN 3059-0442

Journal of Virtual Convergence Research

Volume 1 Nomber 1 Jan.2024

Received: 30 October 2024. Accepted: 01 December 2024

Journal of Virtual Convergence Research
2025, VOL. 1, NO. 1, 55~77
https:// intlmetaverse.com/home/journal?lang=en

International
Metavese
Association

AI-Based Sound Cue Visualization Prototype Study for Improving Interaction Experience of Hearing-impaired in VR Environment

Haneol Kim[1,*], Hyojung Kim[1]

[1] Dept. of Metaverse Graduate School, Sogang Univ., Korea

[*] Corresponding author: Haneol Kim. Email: loenahmik@gmail.com

Abstract

In response to the accessibility challenges faced by hearing-impaired individuals in virtual reality (VR) environments, this study proposes an AI-based sound cue visualization system to enhance their interaction experience. The advancement of XR technologies has revolutionized immersive experiences, yet hearing-impaired users continue to face significant constraints in both verbal and non-verbal communication. Addressing these limitations, this research develops a system leveraging the YAMNet model, an audio event classification model capable of distinguishing between 521 audio classes, to detect and visualize auditory information in real-time. At the core of the system architecture lie three main modules: sound detection and classification, visualization transformation, and VR interface. Through these modules, we implement three distinct prototypes: spatial experience support recognizing environmental and animal sounds to enhance immersion; emotion recognition activation analyzing vocal characteristics to improve communication abilities; and safety and emergency response for detecting and x-alerting environmental hazards. Built on Unity with Meta Quest 3 hardware, the prototype utilizes Sentis for efficient on-device deep learning inference. To evaluate the system's effectiveness, this study proposes comprehensive user experience assessments targeting hearing-impaired individuals and their support personnel. These evaluations will measure VR environment immersion, information comprehension, and satisfaction using a 5-point Likert scale. While representing a significant step toward digital inclusivity, this research requires further empirical testing with hearing-impaired individuals and stakeholders. Future development paths include conducting real user tests in multiplayer VR environments, advancing sound recognition capabilities, and exploring sign language-to-text conversion systems, ultimately contributing to an equitable digital environment where universal participation becomes reality.

Keywords : AI, AI-based sound, VR, hearing-impaired, sound cue visualization

AI-Based Sound Cue Visualization Prototype Study for Improving Interaction Experience of Hearing-impaired in VR Environment

1. Introduction

XR (eXtended Reality) encompasses technologies such as VR, AR, and MR, integrating immersive technology into digital content to create experiences akin to reality. With the commercialization of 5G technology, alongside improved performance and increased distribution of related devices, users can now enjoy more lifelike XR content. 5G technology, with its up to 20Gbps data transmission speed and ultra-low latency of around 1ms, is expected to allow for the implementation of XR content with enhanced immersion and realism. Based on these technological advancements, research and development in the XR field are being actively pursued internationally. Furthermore, XR technology is emerging as a key driver of industry innovation and new business models and markets by integrating with various sectors, showcasing its role as a catalyst for broad economic and social changes(Lee K., 2020).

Due to the advancement of XR technologies, many users can now experience various immersive realities such as virtual, augmented, and mixed reality. However, these innovative technologies emerging in the market are not providing equal experiences to all users. In particular, hearing-impaired individuals face significant constraints in both verbal and non-verbal communication within XR environments.

Existing technologies like OpenAI's Whisper(Radford, 2022), a speech-to-text technology, have made significant progress in supporting verbal communication for the hearing-impaired by converting speech into text. This technology enables hearing-impaired individuals to understand the content of conversations. However, this approach has limitations in capturing non-verbal communication elements. For example, the tone, emotional nuances, and ambient sounds are crucial for understanding the context of conversations, but simple text conversion is insufficient to convey this information.

Therefore, this study aims to address these issues by using AI technology to visualize non-verbal auditory information in VR environments, thus providing a more inclusive and enriched communication experience for the hearing-impaired. The design of AI-based sound cue visualization moves beyond mere verbal communication, enabling holistic communication that includes emotions, intentions, and environmental contexts, and proposes enabling hearing-impaired individuals to have rich experiences in VR environments.

2. Theoretical Background and Previous Research

2.1. The Necessity for Improved Accessibility for the Hearing-impaired

As of the end of May 2023, there are approximately 2.647 million registered disabled individuals in South Korea, of which 432,854 are hearing-impaired. The majority of disabilities, 88.1%, are acquired due to conditions(58.1%), rather than accidents(29.9%). Additionally, the proportion of single-person households among disabled individuals is 26.6%, and there has been an increase in the level of education with 17.4% holding a college degree or higher in 2023, up from 14.4% in 2020. The usage rates of ICT devices have also increased, with 90.5% using mobile phones or smartphones, 29.6% using computers, and 47.7% using the internet, up from 36.5% in 2020.

According to the "Disability Life Panel Survey" of 2022, there are 251,277 disabled individuals using social networking sites(SNS), with hearing-impaired individuals accounting for 73,818. The reasons for using SNS include sharing knowledge and information(49.1%), sharing personal daily life and interests(20.7%), curiosity or seeking new fun and enjoyment(9.7%), and making friends without prejudice against disabilities(6.4%).

Surveys on the utilization and needs of welfare services for the hearing-impaired reveal that the most assistance during social activities comes from acquaintances who are hearing-impaired and sign language interpreters. The major difficulties faced during participation include lack of communication support(34.5%), lack of information about activities(31.3%), not knowing how to participate(14.0%), financial burdens(13.3%), and other factors(6.8%). These findings highlight the need for improved linkage of human and material resources to support communication and enhance accessibility to information for smooth social participation.

The overall educational level of disabled individuals has been rising, necessitating diverse support that aligns with this trend. Although the use of ICT devices has increased, it remains below the levels of the general population, indicating a continued need for efforts to enhance digital accessibility for the disabled. The notable use of SNS by hearing-impaired individuals suggests that SNS plays a significant role as an alternative means of communication. The high demand for information sharing and social relationship formation via SNS indicates the need for the development of disability support services utilizing these platforms. Additionally, the lack of communication support is identified as a major barrier to the social participation of hearing-impaired individuals, indicating the need for an expansion of sign language services and a stronger communication support system.

Furthermore, there are significant gaps in opportunities for communication with family, education, social participation, and the pursuit of happiness due to the lack of sign language education platforms for farming families, inadequate accessibility of sign language books and educational materials, work environments, and various IT devices, as well as insufficient accessibility for cultural enjoyment(Lee, 2022). Thus, there are inequalities in various opportunities throughout the life cycle of hearing-impaired individuals, which can be partially addressed through various accessibility solutions such as sign language translation technologies.

2.2. Accessibility Challenges and Improvement Efforts for the Hearing-impaired in VR

Researchers have been proposing various innovative approaches to enhance AR and VR accessibility for the hearing-impaired. Mauro conducted research to implement visual signals and subtitle systems in virtual environments to improve accessibility for the hearing-impaired in live theaters(Mauro et al., 2018). Kepp developed a "Virtual Reality pitch ranking" system for children using cochlear implants and hearing aids to enhance their spatial hearing skills(Kepp et al., 2022). Vickers demonstrated the importance of user-centered design by involving hearing-impaired children and adolescents in the development of the "BEARS (Both EARS) Virtual Reality Training Suite."(Vickers et al., 2021). Enriquez created a VR application called "SENSE," which utilizes visual and tactile feedback to enhance the musical experience for the hearing-impaired(Enriquez et al., 2020). These diverse approaches contribute to improving the AR/VR experiences of the hearing-impaired and creating more inclusive immersive environments.

2.3. Non-Verbal Communication and AI

The integration of non-verbal communication and artificial intelligence (AI) is emerging as an important research area that enhances the quality of human-machine interaction. Admoni and Scassellati suggested that integrating non-verbal behaviors in interactions with robots could significantly improve communication fluency and efficiency(Admoni and Scassellati, 2015). Building on this concept, Costea developed a real-time non-verbal chatting system that analyzes facial expressions and head movements, showcasing new possibilities for human-AI interaction(Costea et al., 2024). In the medical field, Mohd Azmi emphasized the importance of non-verbal communication in online medical education and studied the impact of body posture, facial expressions, and voice intonation on the learning environment(Mohd Azmi et al., 2024). Furthermore, Kristensson developed AI-based assistive technologies for non-verbal individuals with motor disabilities, attempting an innovative approach that minimizes the physical inputs required for communication(Kristensson et al., 2020). These diverse research approaches illustrate how the integration of non-verbal communication and AI can shape the future of human-machine interactions, suggesting the potential for more natural and effective interaction systems in the future.

3. System Proposal

3.1. Overview of AI-Based Sound Cue Visualization

The proposed system consists of three main modules: a sound detection and classification module, a visualization transformation module, and a VR interface module. The sound detection and classification module utilizes the YAMNet model, an audio event classification model developed by Google, capable of distinguishing between 521 audio classes. This model is based on the MobileNet v1 architecture, making it lightweight and capable of real-time processing on VR devices. The visualization transformation module represents the audio classification results detected and categorized by YAMNet in a probabilistic order of likelihood visually. The VR interface module seamlessly integrates the generated visual cues into the user's VR environment through a panel UI.

3.2. System Environment

The prototype environment was developed based on Unity 6(version 6000.0.20f1), which provides a robust VR framework. The Meta Quest 3 was used as the VR hardware to implement a high-quality immersive experience. Deep learning inference was performed using Sentis version 2.1.0, ensuring efficient on-device processing.

The YAMNet model, trained on Google's Audio Set, was used for acquiring and classifying sounds. Audio Set is a comprehensive dataset that encompasses acoustic events possible in VR environments, consisting of over 2 million sound clips of 10 seconds each, from YouTube content. These audio samples are categorized into 527

Figure 1. *Screenshot of a program that applies audio classification in VR*

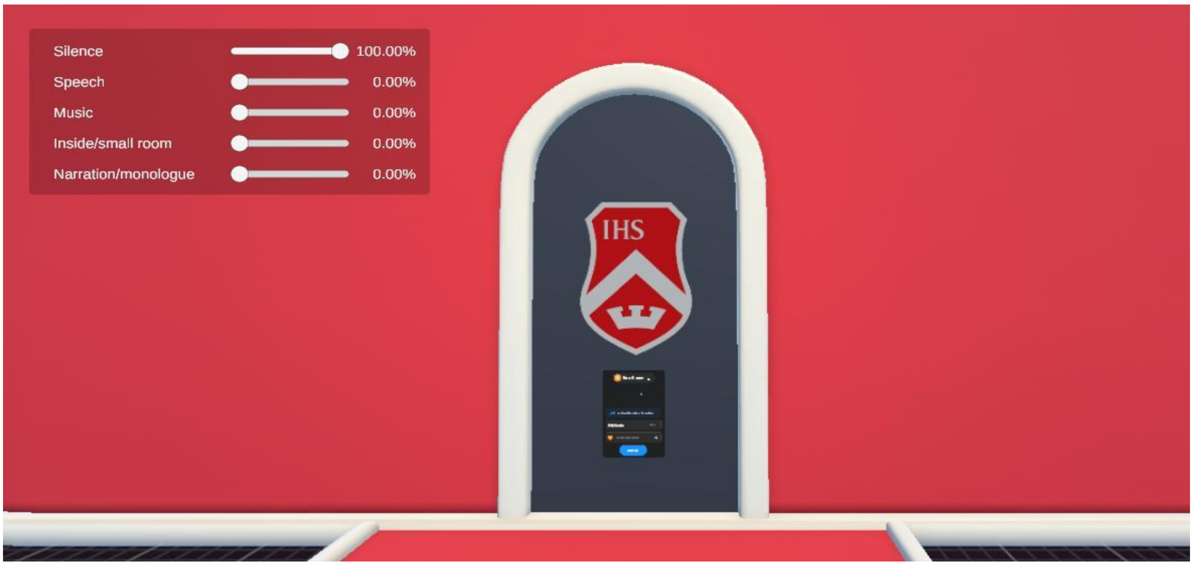
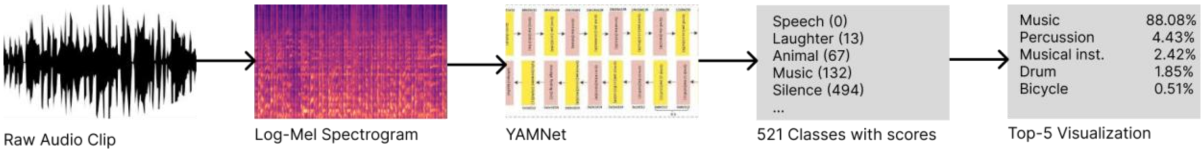


Figure 2. *The pipeline for detecting audio signals into 521 classes*



unique audio event classes, enabling precise sound analysis.

For real-time audio processing and classification, signals inputted through microphones in the VR environment were processed using Sentis. Specifically, classification of the current audio stream was performed every 0.96 seconds. This interval is optimized for real-time processing and matches the YAMNet model's basic input length of 0.96 seconds(16kHz sampling rate with 15,360 samples).

In terms of the user interface(UI) configuration, the top 5 classes with the highest probability among the classified audio events for each 0.96-second segment were visually represented. This approach provides a comprehensive yet intuitive understanding of the current sound environment while minimizing the impact of potential misclassification. The UI clearly displays the name and probability of each class, designed to allow users to immediately grasp the current acoustic environment.

This experimental setup was meticulously designed to effectively detect and classify various auditory stimuli within the VR environment. Through real-time audio classification and intuitive UI representation, the research's primary objective of developing a sound cue visualization system for hearing-impaired individuals is optimized, ensuring enhanced accessibility in VR applications.

3.3. VR Environment Experience Prototype

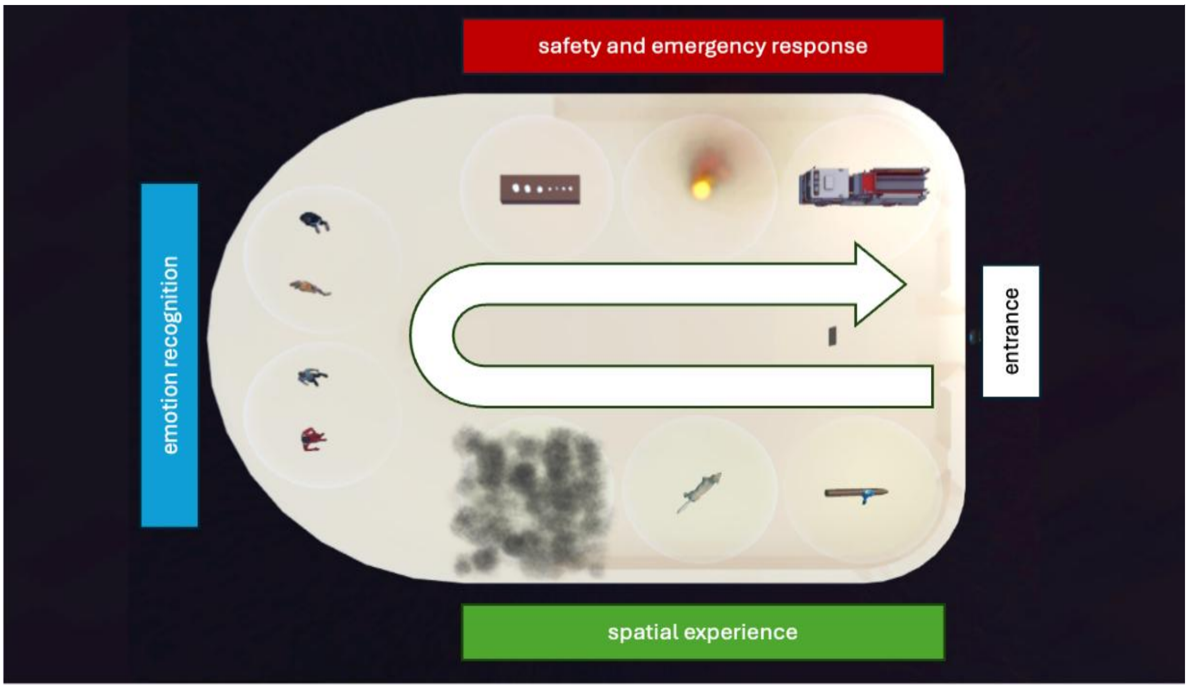
In this study, an AI-based sound cue visualization system utilizing the YAMNet model was applied to various scenarios to enhance the VR experience for hearing-impaired individuals. This system aims to fully leverage YAMNet's ability to classify 521 audio classes to implement customized visualization methods suitable for each scenario.

Thus, the study has developed three types of prototypes considering the situations that hearing-impaired individuals may encounter when entering the VR environment. These prototypes are designed to improve accessibility and enhance the interaction experience for hearing-impaired individuals within the VR environment.

3.3.1. Spatial Experience Support

The first prototype supports an environmental and animal sound recognition system. This prototype visualizes the auditory elements of natural environments to enhance immersion in the VR environment and provide a richer virtual experience for hearing-impaired users. Specifically, it detects natural environmental sounds such as wind, rain, and thunder in real-time and visually represents them. For example, it identifies and categorizes various animal sounds and rain sounds and visually depicts them. This is expected to increase the realism and immersion in the VR environment, improve the quality of the natural environment experience, and enhance the

Figure 3. *Overview of the experience prototype*



environmental awareness abilities of hearing-impaired users.

3.3.2. Emotion Recognition Activation Support

The second prototype supports emotion recognition activation, aimed at assisting in conversation participation and enhancing emotional conveyance to improve the communication abilities of hearing-impaired individuals. This prototype detects nearby conversations, visualizes the location of speakers, and graphically represents the intensity and tone of the conversation to help users understand the flow of dialogue. Additionally, it analyzes the tone, pitch, and volume of voices to infer the emotional state of speakers and visualizes this information to make it perceivable to hearing-impaired users. Through these features, the prototype is expected to enhance the participation of hearing-impaired individuals in conversations, increase their understanding of non-verbal communication, and ultimately improve the quality of social interactions.

Table 1. *Examples of key audio classes used for supporting spatial experiences.*

Situation Classification	Classification Number	Situation Description
Wind	277	Detecting wind sound
Rain	283	Detecting rain sound
Thunder	281	Detecting thunder sound
Animal	67	Detecting animal sound
Bird vocalization	107	Detecting bird sound
Water	282	Detecting water sound
Rustling leaves	279	Detecting rustling leaves
Dog	69	Detecting barking dog sound
Cat	76	Detecting cat meowing sound
Insect	106	Detecting insect sound

Table 2. *Example of a prototype supporting spatial experiences*

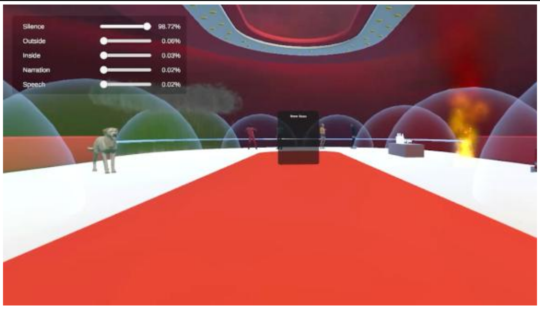
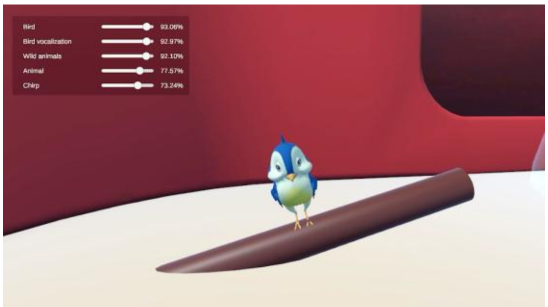
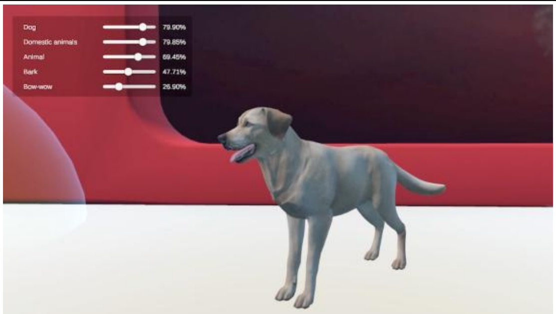
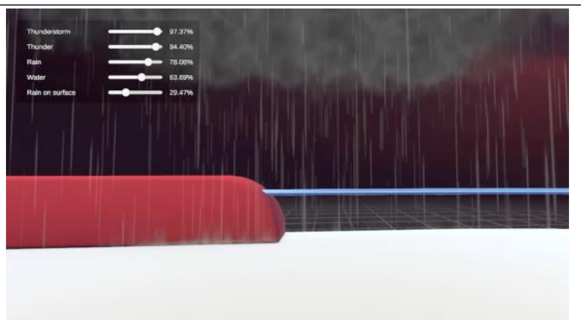
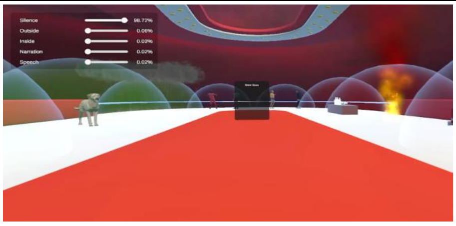
<table>
<tr><th>Prototype 1</th><th>Implementation Results</th></tr>
<tr><td>Hearing-impaired individual "A" enters a virtual space composed of various natural environments with their avatar</td><td>
</td></tr>
<tr><td>↓</td><td></td></tr>
<tr><td>"A" slowly experiences the virtual space composed of natural environments and then perceives that bird sounds are present</td><td>
</td></tr>
<tr><td>↓</td><td></td></tr>
<tr><td>"A" slowly experiences the virtual space composed of natural environments and then perceives that a dog barking sound is present</td><td>
</td></tr>
<tr><td>↓</td><td></td></tr>
<tr><td>"A" slowly experiences the virtual space composed of natural environments and then perceives that rain sounds and thunder sounds are present</td><td>
</td></tr>
</table>

Table 3. *Examples of key audio classes used for supporting emotion recognition*

Situation Classification	Classification Number	Situation Description
Speech	0	Speech recognition and analysis
Conversation	2	Conversation pattern detection
Chatter	63	Detection of surrounding conversations
Yell	9	Detection of loud sounds
Whispering	12	Detection of whispers
Laughter	13	Detection of laughter
Crying	19	Detection of crying
Sigh	23	Detection of sighs
Crowd	64	Detection of crowd noise
Hubbub	65	Detection of background speech noise

Table 4. *Example of a prototype supporting emotion recognition*

Prototype 2	Implementation Results
Hearing-impaired individual “A” enters a virtual space composed of various people with their avatar	
↓	
“A” meets people in the virtual space composed of various individuals and, upon recognizing that laughter is present, shares emotions with other users	
↓	
“A” meets people in the virtual space composed of various individuals and, upon perceiving the sound of people arguing, shares emotions with other users	

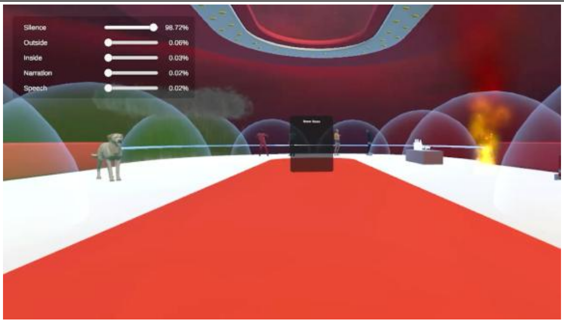
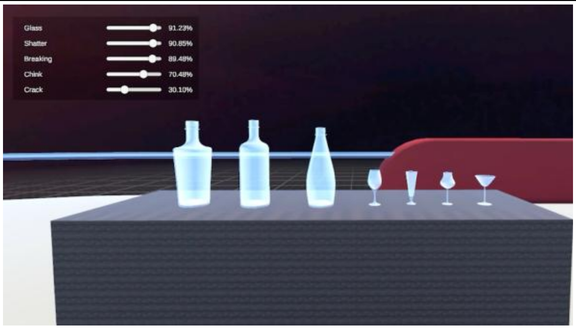
3.3.3. Safety and Emergency Response Support

The third prototype focuses on alerting environmental hazards and emergencies, enhancing safety of hearing-impaired individuals by detecting risks. This prototype has been designed as an educational program on real-world dangers and is implemented to prepare users and hearing-impaired individuals for emergency situations. It prioritizes risk detection and converts them into visual signals to prompt rapid responses. Through this, it is expected to strengthen the safety of hearing-impaired individuals, improve their ability to cope with hazards, and minimize damages through swift responses. Furthermore, plans are in place to enhance this system to address risks in VR environments, such as cyberbullying, by representing them through sounds like car horns, sirens, and breaking glass, and providing visualization to support immediate alerts.

Table 5. *Examples of key audio classes used for supporting safety and emergency response*

Situation Classification	Classification Number	Situation Description
Vehicle horn	302	Detection of car horn
Siren	390	Detection of siren sound
Glass	435	Detection of glass breaking sound
Fire alarm	395	Detection of fire alarm
Civil defense siren	391	Detection of emergency siren
Smoke detector	393	Detection of fire detector alarm
Police car	317	Detection of police siren
Ambulance	318	Detection of ambulance siren
Fire engine	319	Detection of fire truck siren
Alarm	382	Detection of general alarm sound

Table 6. *Example of a prototype supporting safety and emergency situations*

Prototype 3	Implementation Results
Hearing-impaired individual "A" enters a virtual space composed of various safety and emergency situations with their avatar	
↓	
"A" perceives the sound of glass breaking in the virtual space and then responds accordingly	
↓	
"A" perceives the sound of a fire alarm in the virtual space and then responds accordingly	
↓	
"A" perceives the sound of an ambulance in the virtual space and then responds accordingly	

Through the implementation of various scenarios, this study's AI-based sound cue visualization system overall enhances the VR experience for hearing-impaired individuals, enabling better accessibility and participation in a range of situations from everyday life to special experiences.

3.4. Effectiveness Evaluation Plan

This study intends to evaluate the effectiveness of the developed prototypes by conducting user experience assessments targeting hearing-impaired individuals and their support personnel. The evaluation will primarily focus on user experience, measuring immersion in the VR environment, comprehension of information, and satisfaction using a 5-point Likert scale. Additionally, specific user experience items and open-ended surveys will be designed for each prototype-supported situation to verify their effectiveness. Thus, it will be possible to determine how significantly the immersion in the VR environment, comprehension of information, and overall satisfaction are enhanced, and to what extent participation in communication scenarios and understanding of emotions are improved.

Table 7. *Examples of effectiveness evaluation items*

Category	Details
Support for communication and emotion recognition	Conversation participation, understanding of conversation content, satisfaction with emotional expression
Safety and emergency response	Sense of safety, risk perception, emergency response
Prototype Experience	Experience immersion, experience satisfaction, intention to participate in future activities.
Environmental awareness and immersion	Sense of environmental presence, satisfaction with sound recognition, understanding of the ecosystem
System usability evaluation	Ease of use, visualization effectiveness, need for improvement

The effectiveness of the proposed AI-based sound cue visualization system will be assessed through items like those in Table 7, anticipating that this system will enhance the experience of hearing-impaired individuals in VR environments, improve accessibility to VR technologies, and contribute to the digital inclusion of hearing-impaired individuals.

4. Conclusion and Future Research

This study proposed an AI-based sound cue visualization system to enhance the interaction experience of hearing-impaired individuals in VR environments and designed experiments to assess its effectiveness. The system, utilizing the YAMNet model, is expected to effectively detect and visualize audio events in VR, providing hearing-impaired users with a richer and more accessible experience.

The proposed prototypes are expected to significantly improve user experiences in various scenarios, including communication and emotion recognition support, safety response, enhancement of entertainment experiences, and increased environmental awareness. In particular, immersion in VR, understanding of information, and overall satisfaction are expected to improve to a statistically significant extent.

The AI-based sound visualization prototypes proposed in this study have the potential to enhance the experience of hearing-impaired individuals in VR environments. This represents an important advancement towards a more inclusive digital future, with further research likely to expand the capabilities of this technology.

However, this study is based on prototypes from previous research, and empirical testing is necessary. Experiments involving hearing-impaired individuals and stakeholders are needed to collect and analyze data related to the sound visualization system, and to assess improvements in immersion, information understanding, and satisfaction.

This proposal allows for the research findings to be applied in industry. To assess the effectiveness of the proposed system, real user tests in a multiplayer VR environment involving both hearing-impaired and non-impaired individuals are necessary. This would allow for analysis of the system's practicality and user experience.

Additionally, research on an advanced sound recognition system that informs users about the classification, location, and intensity of sounds is necessary. Such a system would deliver sound information within VR to hearing-impaired users accurately, providing a richer experience.

Furthermore, research is needed to develop a translation system that converts sign language to text and vice versa, facilitating communication between hearing-impaired and non-impaired individuals. This would enable natural communication in VR, enhancing user experience.

Such future research would enhance accessibility for disabled individuals in VR and contribute to creating a digital environment where all users can equally participate. This goes beyond technology to promote social inclusion and equality, signifying the importance of this research.

Funding: This work was supported in part by the MSIT(Ministry of Science and ICT), Korea, under the Graduate School of Metaverse Convergence support program (202439002.01-RS-2022-00156318) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

References

Admoni, H., & Scassellati, B. (2015). *Robot Nonverbal Communication as an AI Problem (and Solution)*. AAAI Spring Symposium Series.

Costea, D., Barsan, A., Dimitriadis, D., & Iosif, E. (2024). *Maia: A Real-time Non-Verbal Chat for Human-AI Interaction*. arXiv preprint arXiv:2402.06385.

Creed, C., Al-Kalbani, M., Theil, A., & Metatla, O. (2024). Inclusive AR/VR: accessibility barriers for immersive technologies. *Universal Access in the Information Society*, *23*, 59-73.

Enriquez, K., Palacios, M., Pallo, D., & Guerrero, G. (2020). SENSE: Sensory component VR application for hearing-impaired people to enhance the music experience. *Proceedings of the 2020 15th Iberian Conference on Information Systems and Technologies*, 1-6.

Gemmeke, J. F., Ellis, D. P., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M., & Ritter, M. (2017). Audio Set: An ontology and human-labeled dataset for audio events. *Proceeding of 2017 IEEE International Conference on Acoustics*, 776-780.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. arXiv preprint arXiv:1704.04861.

Kepp, N. E., Arrieta, I., Schiøth, C., & Percy-Smith, L. (2022). Virtual Reality pitch ranking in children with cochlear implants, hearing aids or normal hearing. *International Journal of Pediatric Otorhinolaryngology*, *161*, 111241.

Kristensson, P. O., Bi, X., Mortensen, J., Chen, X., Halvey, M., & Waller, A. (2020). A design engineering approach for quantitatively exploring context-aware sentence retrieval for nonspeaking individuals with motor disabilities. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-13.

Lee, I. (2022). Status and Examples of Sign Language Translation Technology for Improving Accessibility for the Hearing-impaired. *Journal of the Korea Society of Computer and Information*, *40*(10), 14-21.

Lee, K. (2020). Trends in XR (eXtended Reality) Convergence Technologies. *Journal of the Korea Institute of Communications and Information Sciences*, *37*(10), 17-22.

Lee, M. (2010). *Survey on the Usage and Needs of Welfare Services for Hearing-impaired People*. Korea Institute for Disability Development.

Mauro, T., Lourenço, A., Postal, J., & Lucena, V. F. (2018). Exploring virtual reality to enable deaf or hard of hearing accessibility in live theaters: A case study. *Proceedings of the International Conference on Universal Access in Human-Computer Interaction*, 273-284.

Ministry of Health and Welfare. (2024). *2023 Disability Status Survey Press Release Reference Material*. Korea Institute for Health and Social Affairs.

Mohd Azmi, N. A., Pek, J. H., Ang, H. G., Lim, S. H., Lau, T. C., & Samarasekera, D. D. (2024). Beyond words: analyzing non-verbal communication techniques in a synchronous online feedback session for communication skills activities. *BMC Medical Education*, *24*(1), 265.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust Speech Recognition via Large-Scale Weak Supervision*. arXiv preprint arXiv:2202.01184.

Vickers, D., Salorio-Corbetto, M., Driver, S., Rocca, C., Levtov, Y., Sum, K., ... & Kitterick, P. T. (2021). Involving Children and Teenagers with Bilateral Cochlear Implants in the Design of the BEARS (Both EARS) Virtual Reality Training Suite Improves Personalization. *Frontiers in Digital Health*, *3*, 759723.

Xu, W., Liang, H. N., Zhang, Z., & Baghaei, N. (2023). *Inclusive AR/VR: Accessibility Barriers for Immersive Technologies*. arXiv preprint arXiv:2304.13465.