SoundBlender: Manipulating Sounds for Accessible Mixed-Reality Awareness

Ruei-Che Chang
rueiche@umich.edu
University of Michigan
Ann Arbor, MI, USA

Chia-Sheng Hung
yoyung0809@gmail.com
University of Michigan
Ann Arbor, MI, USA

Dhruv Jain
profdj@umich.edu
University of Michigan
Ann Arbor, MI, USA

Anhong Guo
anhong@umich.edu
University of Michigan
Ann Arbor, MI, USA

ABSTRACT

Sounds are everywhere, from real-world content to virtual audio presented by hearing devices, which create a mixed-reality soundscape that entails rich but intricate information. However, sounds often overlap and conflict in priorities, which makes them hard to perceive and differentiate. This is exacerbated in mixed-reality settings, where real-world and virtual sounds can conflict with each other. This may exacerbate the awareness of mixed reality for blind people who heavily rely on audio information in their everyday life. To address this, we present a sound rendering framework SoundBlender, consisting of six sound manipulators for users to better organize and manipulate real and virtual sounds across time and space: Ambience Builder, Feature Shifter, Earcon Generator, Prioritizer, Spatializer, and Stylizer. We demonstrate how the sound manipulators can increase mixed-reality awareness through a simulated working environment, and a meeting application.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design theory, concepts and paradigms**; **Mixed / augmented reality**; **Virtual reality**.

KEYWORDS

VR/AR, extended reality, mixed reality, sound, interaction design, accessibility

ACM Reference Format:
Ruei-Che Chang, Chia-Sheng Hung, Dhruv Jain, and Anhong Guo. 2023. SoundBlender: Manipulating Sounds for Accessible Mixed-Reality Awareness. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct), October 29–November 01, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3586182.3615787

1 INTRODUCTION

Mixed-Reality (MR) enables the concurrent experience of the real world (RW) and virtual reality (VR). The blending of RW and VR sounds occurs commonly in everyday life and is even more difficult to perceive and understand when the corresponding visual information is unavailable. Therefore, it causes challenges for blind people who need to rely on much more audio feedback in many everyday scenarios where RW and VR sounds are mixed and overlapped, such as when navigating the screen using screen readers while doing a hybrid virtual meeting where virtual and real people are both talking or when walking on a busy and noisy street with audio guidance of a navigation app. This situation will be further exacerbated in future MR environments as technology continues to integrate auditory displays into the RW, stemming from various activities that require virtual sounds (e.g., virtual meetings, live streaming). However, while the blending of RW and VR in other modalities was widely explored (e.g., visual [2–4, 6, 7, 9, 11] and haptic [1, 5, 10]), the harmonized blending of RW and VR sounds to be effectively delivered to the end users is still under-explored.

We envision a future in which humans can customize their MR soundscape by manipulating sound properties (e.g., pitch, volume, duration), time, and space from real and virtual worlds. We explore the balanced sound awareness in both realities through six sound manipulations along the Reality-Virtuality continuum [8], which may benefit people who heavily rely on sounds such as blind or visually-impaired people (BVI). We incorporate these six techniques as SoundBlender, which can be coordinated, combined, or extended to create harmonized MR soundscapes:

(1) **Ambience Builder**, which controls the ambient sounds or background noises, to shift the sense of presence by manipulating acoustic transparency to occlusive opacity.
(2) **Feature Shifter**, which controls each sound characteristic, such as the volume, pitch, duration, etc.
(3) **Spatializer**, which controls the audio rendering of a sound such as mono, stereo, and 3D spatial locations.
(4) **Stylizer**, which manipulates a sound's fidelity and overall style through different sound filters (e.g., high, low pass, distortion) and style transfer techniques (e.g., robotic, anime).
(5) **Earcon Generator**, which appends earcons on-demand for certain sound events specified by users, and
(6) **Prioritizer**, which presents sound through user-defined priority by manipulating time (e.g., delaying less important sounds) and sound characteristics (e.g., the more important, the higher the volume).

Next, we discuss the demonstration of SoundBlender in different scenarios.

2 DEMONSTRATING SOUNDBLENDER

We demonstrate SoundBlender through one simulated work environment and one meeting web application. The simulated one is implemented in Unity to simulate the RW and VR, which can provide us with the audio stream data and the control of their playback and manipulations in advance; this enables a smoother user experience without delays or errors from sound recognition or rendering.

UIST '23 Adjunct, October 29–November 01, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0096-5/23/10.
https://doi.org/10.1145/3586182.3615787

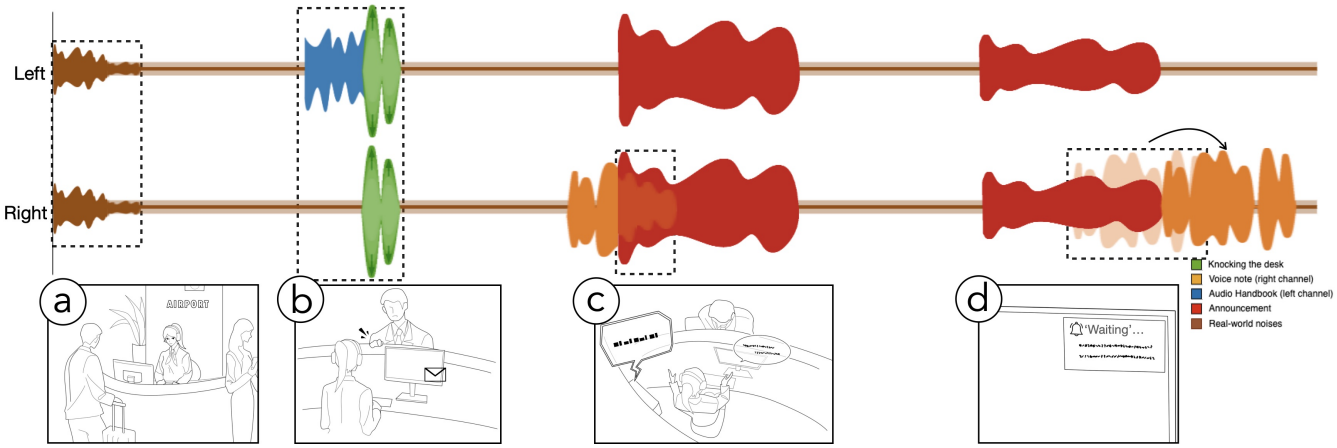

Figure 1: (a) Emma works at the help desk and listens to the audio handbook. She uses Ambience Builder to reduce the RW sounds (brown) to increase her focus. (b) Sometimes people knock on the table to get Emma's attention. She uses Feature Shifter to increase the volume of knocking (from light to dark green). On her monitor, a voice note from her supervisor will be automatically read out on her headphone. She uses Spatializer to place the audio instruction on the left (blue) and the voice note on the right (orange). (c) The volume of voice notes will decrease when being conflict with the announcement. (d) If the voice note is about to conflict with the announcement, Prioritizer will postpone the playback of the voice note to the end of the announcement.

In contrast, the meeting web application can be more practical and familiar for users but may have errors in sound recognition based on the environment.

2.1 Simulated Scenario: Consuming Audio Instructions While Working at the Help Desk

In this scenario, Emma is congenitally blind and has a part-time job at the student center help desk. Her supervisor provided her with the new employee handbook. The handbook is made in an accessible audio version. During her work, people sometimes ask her for help; they knock on the desk to get her attention. During work, Emma's supervisor also sometimes sends voice notes to her; on the other hand, several school announcements require her attention to understand the content.

As having no RW tasks requiring her persistent attention, Emma applies full noise cancellation using Ambience Builder to focus on the handbook (Figure 1a). She then places the handbook on her left using Spatializer, and voice messages on her right for easier distinction. Since the knocking events from RW are suppressed by the Ambience Builder, Emma amplifies them using Feature Shifter to attend to her job duty and be aware of people coming (Figure 1b). She also shifts the level of noise cancellation to half using Ambience Builder for building announcements and amplifies it using Feature Shifter; otherwise, the announcements are sometimes suppressed due to the full noise cancellation.

To address the conflicts between RW and VR sounds, she coordinates the Prioritizer and Feature Shifter to make announcements and knocking sounds louder than her supervisor's voice notes if notes were played first and conflicted with them (Figure 1c). Also, with Prioritizer, the voice messages will be delayed if they happen during the playback of announcements or knocking (Figure 1d).

Demonstration: We will demonstrate our Unity implementation of this scenario and another two VR scenarios (e.g., one about walking on a busy street and one about a future hybrid meeting) along with three sound conditions for each scenario: Full Transparency, Noise Cancellation, and SoundBlender. With Full Transparency, users can hear the RW sounds but may be in conflict with the VR sounds, making them hard to distinguish. With Noise Cancellation, the RW sounds may be slightly/totally disabled but can hear VR sounds more clearly. By comparing these two methods, the SoundBlender method would provide a more complete experience to consume RW and VR without losing much audio information.

2.2 Accessible Meeting Web Application for BVI People

When engaging in a virtual meeting, BVI people typically need to not only follow the conversation of the meeting but also are required to navigate the meeting application if necessary. For instance, BVI people would be asked to press "raise hand" to poll, mute/unmute the audio if necessary, or check chat messages. However, for BVI people to select and use the meeting functions, they need to use a screen reader to navigate the user interface, where a system voice will be played when a button/function is navigated and focused and another audio feedback will be followed up if the button is pressed/selected (Figure 2a). The audio feedback of the screen reader can be in conflict with the meeting conversation, which may cause BVI people to miss either or both information. Furthermore, BVI people may situate in a noisy environment and encounter interference from the real world when doing a meeting. Some real-world events are unimportant and distracting from the meeting (e.g., people chatting nearby), while others may require user attention (e.g., someone knocking on the door and the alarm going off). Next, we explain how we can manipulate sounds to address these problems.

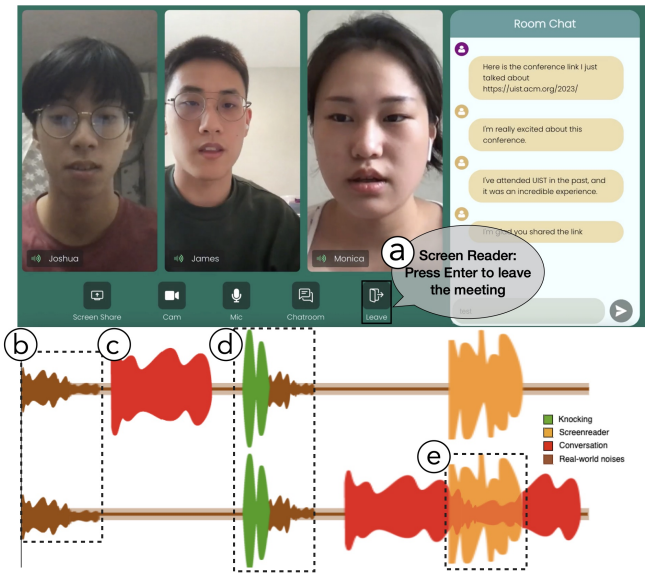

Figure 2: (a) When focusing on the web element, the screen reader will read out the alt-text or other accessibility labels. (b) To focus on the meeting in a noisy environment, the noise cancellation mode is turned on. (c) The voice of the other two people will be assigned as left and right audio channels. (d) When a knocking sound is detected, the noise cancellation will be turned into transparency mode to make the user easier to perceive it. And the mode will switch back to noise cancellation afterward. (e) When using a screen reader to navigate, the volume of the meeting will be decreased to accentuate the audio feedback of the screen reader.

2.2.1 Auto-Switching Ambiance to observe RW events and engage in virtual meeting. Suppose the meeting is taking place in a coffee shop where there are different kinds of sound events, a BVI user may like to reduce the irrelevant noises as possible, so they turn on the noise cancellation mode on their headphone (Figure 2b). However, this could also block out some important sounds at the same time. To address this, in our application, there is a sound recognition module that can continuously detect sound events in the background. The user can specify a list of important real-world sounds to the system. Once the important sounds are detected, the noise cancellation mode will be gradually shifted to the transparency mode and the sound of the meeting will be decreased to allow the user to observe the sound events more easily (Figure 2d). Conversely, when the important sound disappears, the system will switch back to the noise cancellation mode to increase the engagement of the virtual meeting. The detection of real-world sounds and the control of ambiance is enabled by the Ambience Builder in SoundBlender.

2.2.2 Auto-Spatializing Interface Layout and User Voices for Faster Navigation and Immersion. To make the virtual meeting more immersive, SoundBlender in the meeting application spatializes the voices in 3D space to simulate a real meeting for a better sense of immersion. Furthermore, during the meeting, it is common to do actions specifically for a person, such as sending a private message, sending emoji, or mute/unmute. It is thus to be helpful to help BVI users quickly navigate and find the person they want. We thus also dynamically change the user layout to map to the 3D location of attendees' voices so that BVI people can intuitively and quickly navigate the screen reader to the location of the voice (Figure 2c,e).

2.2.3 Harmonizing Different Streams of Virtual Sounds. To display immediate feedback on the screenreader to the user, SoundBlender can harmonize the screenreader and the conversation by controlling the sound characteristics of the screenreader and meeting conversations. For example, Prioritizer controls the audio streams and will outstand the stream of the screen reader by increasing its volume while decreasing the volume of meeting attendees (Figure 2e). We also enable other characteristics such as pitch, voice gender, voice font, speed of speech, etc, so that the user can customize based on their preferences. Conversely, if the meeting is more important, the user can also prioritize the voice of attendees.

Demonstration We will demonstrate the audio features of our meeting web application with multiple users. People can wear headphones and navigate the screen with a screen reader to consume the full audio experience we described above.

3 FUTURE WORK

We plan to develop more simulated scenarios under the SoundBlender framework and conduct studies with BVI and sighted people to understand users' preferences and the timings to use certain manipulators. We will also develop more practical applications on existing platforms to demonstrate the feasibility of SoundBlender on a wide range of domains.

REFERENCES

[1] Florian Daiber, Donald Degraen, André Zenner, Tanja Döring, Frank Steinicke, Oscar Javier Ariza Nunez, and Adalberto L. Simeone. 2021. Everyday Proxy Objects for Virtual Reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 101, 6 pages. https://doi.org/10.1145/3411763.3441343

[2] Isamu Endo, Kazuki Takashima, Maakito Inoue, Kazuyuki Fujita, Kiyoshi Kiyokawa, and Yoshifumi Kitamura. 2021. ModularHMD: A Reconfigurable Mobile Head-Mounted Display Enabling Ad-Hoc Peripheral Interactions with the Real World. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 100–117. https://doi.org/10.1145/3472749.3474738

[3] Uwe Gruenefeld, Jonas Auda, Florian Mathis, Stefan Schneegass, Mohamed Khamis, Jan Gugenheimer, and Sven Mayer. 2022. VRception: Rapid Prototyping of Cross-Reality Systems in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 611, 15 pages. https://doi.org/10.1145/3491102.3501821

[4] Jeremy Hartmann, Christian Holz, Eyal Ofek, and Andrew D. Wilson. 2019. RealityCheck: Blending Virtual Environments with Situated Physical Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300577

[5] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1957–1967. https://doi.org/10.1145/2858036.2858134

[6] David Lindlbauer and Andy D. Wilson. 2018. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173703

[7] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in*

Computing Systems (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2143–2152. https://doi.org/10.1145/2702123.2702382

[8] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[9] Joan Sol Roo and Martin Hachet. 2017. One Reality: Augmenting How the Physical World is Experienced by Combining Multiple Mixed Reality Modalities. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 787–795. https://doi.org/10.1145/3126594.3126638

[10] Adalberto L. Simeone, Eduardo Velloso, and Hans Gellersen. 2015. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3307–3316. https://doi.org/10.1145/2702123.2702389

[11] Chiu-Hsuan Wang, Bing-Yu Chen, and Liwei Chan. 2022. RealityLens: A User Interface for Blending Customized Physical World View into Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 49, 11 pages. https://doi.org/10.1145/3526113.3545686