UltrasonicSpheres: Localized, Multi-Channel Sound Spheres Using Off-the-Shelf Speakers and Earables

Michael Küttner*
michael.kuettner@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Valeria Zitz*
valeria.zitz@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Kathrin Gerling
kathrin.gerling@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Michael Beigl
michael.beigl@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Tobias Röddiger
tobias.roeddiger@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

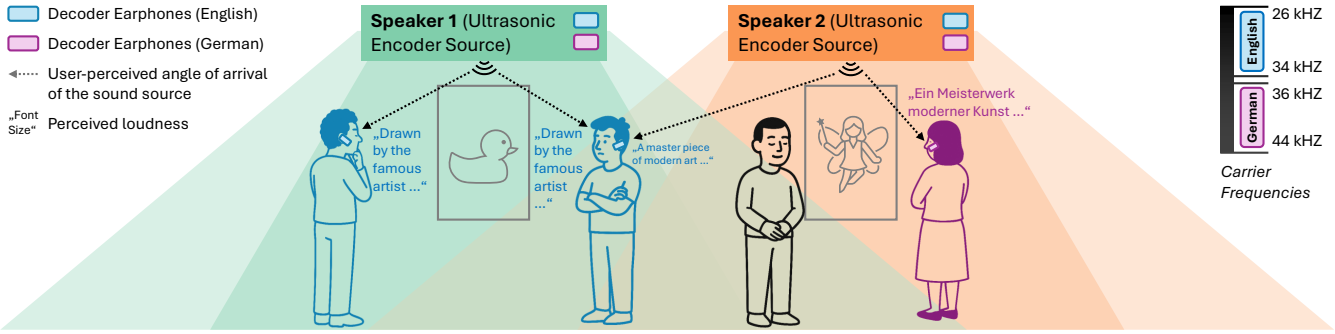

Figure 1: UltrasonicSpheres enables location-specific audio delivery using ultrasonic sources. The system transmits inaudible ultrasound that is converted by users' earphones into audible audio. This preserves spatial awareness and loudness perception, as if the sound was emitted audibly by the location-specific speaker. By encoding different audio channels (e.g., languages) on separate ultrasonic bands, UltrasonicSpheres allows multiple streams to be broadcast simultaneously from a single source. Users can then select their preferred stream directly on the earable device. Users without the corresponding earphones remain unaffected as they cannot hear the emitted ultrasound.

Abstract

We present a demo of UltrasonicSpheres, a novel system for location-specific audio delivery using wearable earphones that decode ultrasonic signals into audible sound. Unlike conventional beamforming setups, UltrasonicSpheres relies on a single ultrasonic speaker to broadcast localized audio with multiple channels, each encoded on a distinct ultrasonic carrier frequency. Users wearing our acoustically transparent earphones can demodulate their selected stream, such as exhibit narrations in a chosen language, while remaining fully aware of ambient environmental sounds. The experience preserves spatial audio perception, giving the impression that the sound originates directly from the physical location of the source. This enables personalized, localized audio without requiring pairing, tracking, or additional infrastructure. Importantly, visitors not equipped with the earphones are unaffected, as the ultrasonic signals are inaudible to the human ear. Our demo invites participants to explore multiple audio zones and experience how UltrasonicSpheres supports unobtrusive delivery of personalized sound in public spaces.

*Both authors contributed equally to this research.

Keywords

earables, hearables, ultrasound, public spaces, earphones

1 Introduction

"Sound zones" enable multiple co-located listeners to experience different audio content simultaneously without cross-interference [2]. These zones are especially useful in environments such as museums, libraries, and open offices, where personalized or context-aware audio enhances user experience without disturbing others. One promising approach to implementing sound zones involves the use of ultrasonic audio beams to project sound into precisely defined spatial regions. Commercial solutions like the Holosonic Audio Spotlight use modulated ultrasound to produce highly directional, laser-like audio beams [6]. However, such systems are expensive and proprietary. Research alternatives include phased arrays of ultrasonic transducers [9] and acoustic metasurfaces [13], which offer precise control over sound localization. Despite their capabilities, these approaches require complex hardware, such as large transducer arrays or custom-fabricated components, limiting their accessibility and scalability.

In this demo paper, we introduce UltrasonicSpheres, a low-cost system that creates spatially confined "sound spheres" using commodity speakers and earphones. The core concept of UltrasonicSpheres is to transmit audio modulated onto inaudible ultrasonic carriers, which propagate through the air and are demodulated into audible sound by the earphone hardware. The system first modulates the desired audio content onto ultrasonic signals, which are emitted by strategically placed speakers. Compatible earphones can then demodulate these signals into clear, audible audio. Since the signal is independently demodulated in each ear, the interaural time difference (ITD) is preserved, thus perceptually retaining the source's location. Furthermore, UltrasonicSpheres support multiple concurrent channels by assigning each channel to a separate ultrasonic carrier frequency. For instance, this allows the delivery of different languages within the same physical zone. Because ultrasonic signals exhibit strong spatial directivity, users can transition between different UltrasonicSpheres zones simply by moving through space. As they enter a sound sphere, their earphones demodulate the corresponding ultrasonic signal and play the associated audio; as they move away, the signal fades and another zone can take over. This enables an intuitive, interface-free method of audio selection based on physical location and orientation.

Attendees of the demo will be able to spatially experience different localized UltrasonicSpheres of two museum-like exhibits. They will experience how the system retains spatial perception and allows switching between different channels (e.g., switching between English and German narration).

2 Related Work

2.1 Localized Spatial Audio in Public Spaces

Spatial audio research has developed techniques that confine audible content to small regions. A common method is to use parametric ultrasonic arrays to produce ultrasound beams at high intensities and leveraging the air's non-linearity to demodulate the carrier, producing audible sound only in the beam path [8]. For example, Audio Spotlight [6] is a commercial system that applies the approach to deliver localized audio in museum exhibits. However, parametric speakers and their high-voltage amplifiers are specialized, costly and power-hungry. To add more flexibility, researchers have applied phased-array control to produce multiple highly localized sounds. For example, Ochiai et al.'s Holographic Whisper [9] steers phase delays across an ultrasonic array to focus energy at arbitrary 3D positions. Each focal spot self-demodulates into a millimetre-scale virtual loudspeaker, and many such spots can be toggled independently to whisper different messages to people in the same space. Further, Zhong et al. [13] published "audible enclave". The system uses the intersection of two self-bending beams shaped by 3D-printed metasurface lenses. Neither beam is audible in flight, but where they cross, a nonlinear interaction generates sound, creating a pocket that only the listener can hear from within. While these systems demonstrate precise spatial targeting, they require custom hardware, complex calibration, and fixed installations.

UltrasonicSpheres addresses this limitation by requiring only a single ultrasonic source and demodulating earphones to deliver audio that is perceivable only within a confined spatial region.

2.2 Earables and Ultrasound

In earable research (i.e., research that explores sensing in and around the ear), ultrasound has been used as a probing signal for different applications from ear canal based authentication [5], to mid-air gesture recognition [7], and heart rate tracking [4]. Most closely related to our work, Watanabe et al. [12] have shown that ultrasonic transmitters can be used for malicious interference with earphones by injecting commands into a victim's noise canceling earbud, which demodulates them into audible messages. While this poses a potential security threat, it also shows the ability of commodity ear-worn devices to receive ultrasound audio.

UltrasonicSpheres builds on this capability, but instead transmits purposefully curated audio to earphone receivers for targeted consumption.

2.3 Earphone-Based Sound Spheres

Commercial hearables, such as the *AirPods Pro*, incorporate spatial audio features that simulate immersive, three-dimensional sound environments by dynamically rendering audio relative to the user's head position [1]. Recent research has extended this concept beyond entertainment to enhance situational audio control. For instance, Chen et al. [3] introduced "sound bubbles", allowing users to define personal audio zones with radii ranging from 1 to 2 meters. Within these zones, hearables selectively permit audio transmission, effectively isolating the user from external sound sources outside the defined perimeter. In parallel, Veluri et al. [11] proposed an approach for selective hearing in social settings, where users can focus on a speaker for several seconds to enroll them as a preferred source, enabling the hearable to prioritize their voice while suppressing others. UltrasonicSpheres extends this idea by externalizing spatial audio rendering into the environment. Rather than simulating zones in the earphones, it uses ultrasound to create spatially anchored sound spheres, which are demodulated for position-based access to personalized audio.

3 UltrasonicSpheres: Localized, Multi-Channel Sound Spheres via Ultrasound

UltrasonicSpheres creates localized, user-specific audio delivery using ultrasound and wearable earphones. The system operates

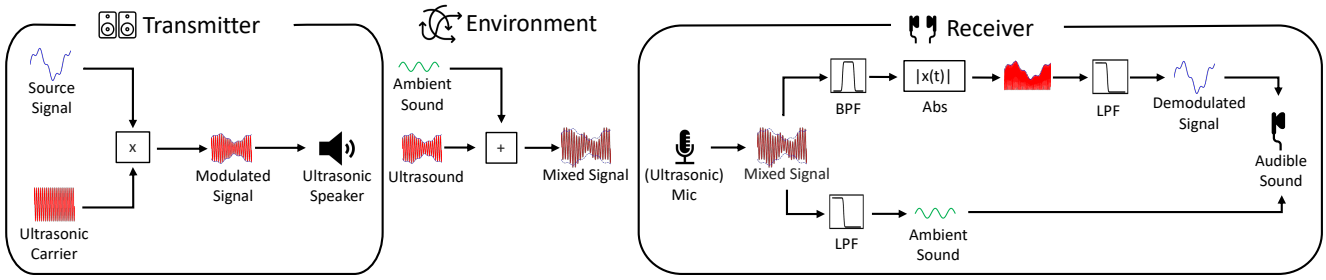

Figure 2: The original signal is modulated and transmitted via a speaker, where it becomes mixed with environmental sound. The receiver captures the combined signal, separates the ultrasonic signal from the noise, and demodulates it. This is achieved by first applying a band-pass filter to isolate the desired ultrasonic channel, followed by a low-pass filter applied to the absolute value of the signal for demodulation into audible sound. Finally, the demodulated signal is reintegrated with the environmental sound to preserve spatial and situational awareness.

entirely radio transmission free: audio signals are modulated using amplitude modulation (AM) with an ultrasonic carrier frequency and are continuously emitted by the speaker. Sound becomes perceptible only to users wearing earables and who are standing inside the acoustic beam. We make use of the fact that high-frequency sound waves travel in a more directed manner to achieve a restriction of sound propagation. Since the signal still travels at the speed of sound, the interaural time difference (ITD) is preserved and the user will naturally perceive the location of the sound source. Using different carrier frequencies, we enable multiple users to receive different content simultaneously in the same physical space, without interference or awareness by others.

The UltrasonicSpheres system consists of two main components: a single-speaker ultrasonic transmitter and two earphone receivers. For the transmitter, we use an off-the-shelf, standard high-fidelity bookshelf speaker readily equipped with a tweeter capable of outputting ultrasonic frequencies (20–55 kHz). We drive it using a Scarlett Focusrite 2i2 at an output sampling rate of 96 kHz, which is connected to an standard high-fidelity amplifier. No special modifications were made to the speakers; the only requirement is that their high-frequency response, along with the driver chain, is capable of handling the ultrasonic carrier frequencies. In our prototype, we found that many consumer audio tweeters, which often extend into the lower ultrasonic range, can serve this purpose. This means that our system can be realized without investing in specialized ultrasonic emitters. For the receiver end, we leverage the open-source OpenEarable 2.0 platform [10], which gives low-level access to the audio processing capabilities of modern noise-canceling earphones. No other special infrastructure is required.

3.1 Modulation

To shift the audio signal into the ultrasonic range, we apply amplitude modulation (AM), a technique well known from radio transmission but equally applicable to acoustic signals. The signal is first low-pass filtered to 4000 Hz to limit bandwidth, then multiplied with a carrier of frequency $f_c = 30\,kHz$.

$$y(t) = [1 + k_a \cdot x(t)] \cdot \cos(2\pi f_c \cdot t), \tag{1}$$

This results in the original signal $X(f)$ being shifted by f_c inside the spectrum, as illustrated in Figure 3. We chose a carrier frequency

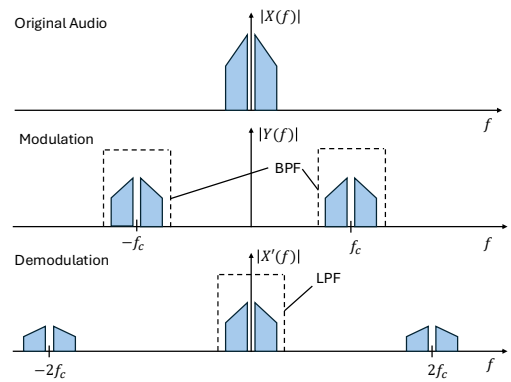

Figure 3: Illustration of the magnitude spectrum in the frequency domain showing modulation and demodulation.

of 30 kHz to ensure that the lower end of the frequency band of the modulated signal $Y(f)$ stays well outside the audible range with 26 kHz.

3.2 Demodulation

To make the modulated ultrasonic signal audible again, we need to demodulate it inside the earphone. To achieve this, the modulated signal is usually multiplied with the carrier signal again and then low-pass filtered afterwards.

$$\begin{aligned} x'(t) =& y(t) \cdot \cos(2\pi f_c \cdot t) \\ =& \frac{1 + k_a \cdot x(t)}{2} + \frac{1 + k_a \cdot x(t)}{2} \cdot \cos(4\pi f_c \cdot t) \\ \hat{x}(t) =& \frac{2 \cdot LPF\{x'(t)\} - 1}{k_a}, \end{aligned} \tag{2}$$

However, this approach requires the receiver to retrieve the phase-synchronized carrier signal. To achieve this, the carrier must typically be recovered from the modulated signal, e.g. using a phase-locked loop (PLL). As a simple alternative, the absolute value of the modulated signal can instead be used. This removes the need for explicit carrier recovery and will still provide sufficient robustness against noise for our use-case.

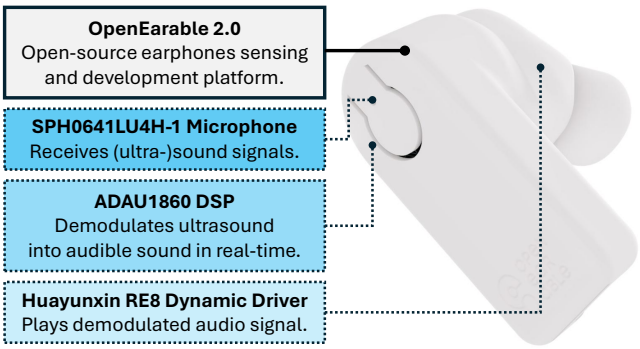

Figure 4: UltrasonicSpheres leverages the open-source OpenEarable 2.0 [10] and its edge audio processing capabilities, resembling those of standard ANC earphones.

The audio signal can be recovered by the following equation (3) using a custom amplification factor A.

$$\tilde{x}(t) = 2A \cdot BPF\{|y(t)|\} \tag{3}$$

3.3 Implementation

On the transmitter side, the audio file is first low-pass filtered and then modulated using MATLAB. To support multiple audio channels, the individual modulated signals are combined into a single composite signal. The result is then exported as a WAV file with a sampling rate of 96 kHz. This file can be simply played back through the connected speaker system. For implementation of the receiver, we adjust the routing and filterbanks of the OpenEarable 2.0's DSP (ADAU1860). To allow proper sampling of ultrasonic signals, a sample rate of 192 kHz was used. The system contains ultrasonic-capable PDM microphones of type Knowles SPH0641LU4H-1 allowing to capture signals of up to 80kHz.

The DSP pipeline comprises three filter stages. The first stage applies a bell-curve equalizer to compensate for imbalances in the ultrasonic frequency response. Next, a Chebyshev Type II band-pass filter isolates the frequency band of interest. This filter type was selected for its steep roll-off and minimal pass-band distortion. As explained above, the absolute value of the signal is then calculated and passed through a low-pass filter to eliminate the high-frequency components. Finally, the signal is amplified and processed through a limiter for dynamic range compression, which helps suppress unwanted clicking sounds from extraneous sources leaking into the ultrasonic band, before being sent to the DAC connected to the dynamic driver. The entire chain is illustrated in Figure 2.

4 Demonstration

We demonstrate UltrasonicSpheres through an exhibit that resembles a typical scenario at a museum (see Figure 5), which lets attendees walk through different localized UltrasonicSpheres, each providing multiple audio channels. The setup includes two ultrasonic speakers, each positioned at a separate exhibit. Both speakers continuously broadcast exhibit-related audio content. One exhibit is accompanied by a soundscape, the other exhibit is supported by related information material on two audio channels: one in English and one in German. The spatially located speakers enable simultaneous, location-specific delivery.

Figure 5: (Left) A visitor experiencing a localized UltrasonicSphere soundscape; (Right) Two visitors looking at an exhibit surrounded by an UltrasonicSphere, each decoding the audio narration to their desired language.

Participants of the demo equipped with demodulation earphones of UltrasonicSpheres can hear the content by simply walking up to the exhibits. They will be able to experience how the ultrasonic signals are turned directly into audible sounds inside the ear canal that can be spatially associated with the location of the exhibit. Further, the demodulation earphones are fully acoustically transparent. This will allow demo participants to remain aware of their surroundings and interact naturally within the exhibition space. Lastly, participants will be able to switch between the German and English channels by tapping the earphones. Demo attendees will also be invited to explore the absence of audible sounds when listening to UltrasonicSpheres without wearing the required earable receivers.

5 Conclusion

We presented UltrasonicSpheres, a lightweight system that enables spatially localized and personalized audio delivery using ultrasound and earable devices. By leveraging off-the-shelf speakers and open-source wearable hardware, UltrasonicSpheres allows users to hear context-specific content, such as ambient soundscapes or exhibit descriptions, without isolating them from their environment, while operating fully standalone and radio transmission free.

Our demonstration shows that selective audio access can be achieved through a combination of spatial orientation and ultrasonic signal decoding, enabling intuitive and unobtrusive interaction. The system supports user-controlled channel switching (e.g., language selection) via simple ear tap gestures, and is well suited for use in public spaces such as exhibitions or galleries.

Acknowledgments

This work has been partly founded by the Carl-Zeiss-Foundation as part of "stay young with robots" (JuBot) project and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK 2739/1 – Project No. 447089431 – Research Training Group: KD2School – Designing Adaptive Systems for Economic Decisions.

References

[1] Apple Inc. 2025. Control Spatial Audio and head tracking. https://support.apple.com/en-en/guide/airpods/dev00eb7e0a3/web Accessed: 2025-05-24.

[2] Terence Betlehem, Wen Zhang, Mark A Poletti, and Thushara D Abhayapala. 2015. Personal sound zones: Delivering interface-free audio to multiple listeners. *IEEE Signal Processing Magazine* 32, 2 (2015), 81–91.

[3] Tuochao Chen, Malek Itani, Sefik Emre Eskimez, Takuya Yoshioka, and Shyamnath Gollakota. 2024. Hearable devices with sound bubbles. *Nature Electronics* (2024), 1–12.

[4] Xiaoran Fan, David Pearl, Richard Howard, Longfei Shangguan, and Trausti Thormundsson. 2023. Apg: Audioplethysmography for cardiac monitoring in hearables. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–15.

[5] Yang Gao, Wei Wang, Vir V Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.

[6] Holosonics. 2025. Audio Spotlight by Holosonics. https://www.holosonics.com/. Accessed: 2025-05-22.

[7] Yincheng Jin, Yang Gao, Yanjun Zhu, Wei Wang, Jiyang Li, Seokmin Choi, Zhangyu Li, Jagmohan Chauhan, Anind K Dey, and Zhanpeng Jin. 2021. SonicASL: An acoustic-based sign language gesture recognizer using earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–30.

[8] Shangming Mei, Hui Xu, Yihua Hu, Mohammed Alkahtani, and Yangang Wang. 2022. The Parametric Array Speaker: A Review. In *International Joint Conference on Energy, Electrical and Power Engineering*. Springer, 1254–1271.

[9] Yoichi Ochiai, Takayuki Hoshi, and Ippei Suzuki. 2017. Holographic whisper: Rendering audible sound spots in three-dimensional space by focusing ultrasonic waves. In *proceedings of the 2017 CHI conference on human factors in computing systems*. 4314–4325.

[10] Tobias Röddiger, Michael Küttner, Philipp Lepold, Tobias King, Dennis Moschina, Oliver Bagge, Joseph A Paradiso, Christopher Clarke, and Michael Beigl. 2025. OpenEarable 2.0: Open-Source Earphone Platform for Physiological Ear Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 1 (2025), 1–33.

[11] Bandhav Veluri, Malek Itani, Tuochao Chen, Takuya Yoshioka, and Shyamnath Gollakota. 2024. Look once to hear: Target speech hearing with noisy examples. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–16.

[12] Hiroki Watanabe and Tsutomu Terada. 2023. UltrasonicWhisper: Ultrasound Can Generate Audible Sound in Your Hearable. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers*. 132–134.

[13] Jia-Xin Zhong, Jun Ji, Xiaoxing Xia, Hyeonu Heo, and Yun Jing. 2025. Audible enclaves crafted by nonlinear self-bending ultrasonic beams. *Proceedings of the National Academy of Sciences* 122, 12 (2025), e2408975122.