Reperceiving Everyday Sounds with Hidden Texture

Yujie Tao
Stanford University
Stanford, California, USA
yjtao@stanford.edu

Mateo Larrea
Stanford University
Stanford, California, USA
mlarreaf@stanford.edu

Portia Wang
Stanford University
Stanford, California, USA
portiaw@stanford.edu

Jingjin Li
AImpower.org
Mountain View, California, USA
jingjin@aimpower.org

Takako Fujioka
Stanford University
Stanford, California, USA
takako@ccrma.stanford.edu

Jane L. E
Computer Science
National University of Singapore
Singapore, Singapore
ejane@nus.edu.sg

Sean Follmer
Stanford University
Stanford, California, USA
sfollmer@stanford.edu

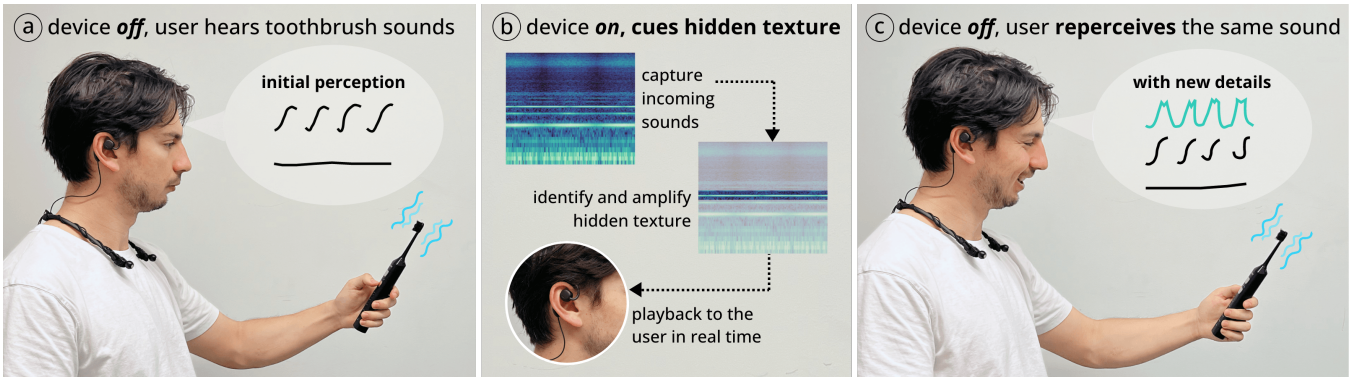


Figure 1: We propose a wearable system that guides users to notice hidden textures in everyday sounds. (a) With the device off, the user's *initial perception* of an everyday sound (e.g., electric toothbrush) is dominated by a few characteristics, as reflected in their drawings. (b) With the device on, the system identifies and amplifies hidden textures in real time, cueing attention to these details. (c) After the device is turned off, the user temporarily notices *new details* when they hear the same sound again (i.e., *reperceive*). Sound drawings in this figure are traced from a participant's drawings in Study #2.

Abstract

Ubiquitous hearable devices often prioritize digital immersion by suppressing real-world sounds, distancing users from their physical surroundings. Conversely, noticing rich acoustic details in everyday sounds brings benefits such as supporting mindfulness and fostering musical appreciation. We explore how hearables can scaffold users to hear "more" of their everyday soundscapes. We present a wearable system that identifies hidden textures in everyday sounds, specifically overtones that shape timbre but are often perceptually hidden. Grounded in prior literature in auditory cueing, the system amplifies these overlooked textures in real time to produce a short-term attentional shift, enabling users to continue noticing these details after device use. We evaluated the system through two studies. A psychoacoustic study (n = 32) showed that the overtone textures detected by our system could be effectively cued. Across 10 sounds, participants showed a 13% improvement in detecting the overtone texture immediately after one-time cueing. These results were further validated by an in-lab user study (n = 12), where participants wore the system while interacting with everyday objects. In 69% of trials, participants reported that cued texture became more perceptually salient after using the device than before. Qualitative accounts further suggested that the device use could encourage participants to form new mental associations with familiar sounds.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

UIST '26, Detroit, MI, USA

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2856-3/26/11
https://doi.org/10.1145/3830398.3830551

CCS Concepts

• **Human-centered computing** → **Sound-based input / output**.

Keywords

Hearables, Sensory Augmentation, Audio Augmented Reality

ACM Reference Format:
Yujie Tao, Mateo Larrea, Portia Wang, Jingjin Li, Takako Fujioka, Jane L. E, and Sean Follmer. 2026. Reperceiving Everyday Sounds with Hidden Texture. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26), November 02–05, 2026, Detroit, MI, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3830398.3830551

1 Introduction

Hearable devices, such as earbuds, have become ubiquitous in everyday life [25], continuously mediating how users perceive their acoustic environment. Noise cancellation remains the dominant operating mode, suppressing environmental sounds to enhance immersion in digital content [54, 69]. While effective for reducing distraction, constantly attenuating real-world sounds distances users from their immediate soundscapes [6, 34, 66].

Being able to notice rich details in everyday sounds can offer many benefits. For instance, sounds from routine activities as simple as tooth brushing or cooking can become anchors for fostering mindfulness, directing user attention back to the the present moment when it drifts [19, 20, 60]. Similarly, attending to natural soundscapes, such as birdsong, can facilitate recovery from stress [14, 55]. Everyday sounds also serve as a rich source of creative inspiration, with musicians blending everyday acoustic details into their compositions [18, 49]. Despite the benefits, everyday listening is typically goal-directed; we tend to focus on functional meanings of sounds (e.g., "a running microwave") rather than appreciate their underlying sonic textures [16]. Even when sounds are physically complex, auditory perception tends to be dominated by a few characteristics, such as primary pitch, leaving finer spectral details to become perceptually fused and remain hidden [22, 51, 52].

Prior work on hearables for supporting awareness of everyday sounds has largely focused on mediating auditory experiences *during* device use—for example, isolating speech in noisy environments [9], selectively filtering environmental sounds using text input [64], or amplifying subtle sounds produced by hand interactions [60]. However, these solutions confine heightened awareness of auditory details to periods of device use: once it is removed, the auditory richness it reveals often becomes inaccessible again.

We explore an alternative role for hearables as temporary scaffolds that redirect users' attention toward overlooked acoustic textures *beyond* device use. Our approach is grounded in auditory cueing [18, 49], using preview cues to guide attention toward target frequency ranges. Specifically, we introduce a wearable system that amplifies and plays back perceptually hidden acoustic textures in real time to induce a short-term attentional shift, enabling users to continue noticing these details after device use. The system continuously monitors the acoustic environment to estimate the primary pitch and identify candidate acoustic components. As an initial investigation, we focus on overtones—tonal components that shape the timbre of everyday sounds but are often perceptually fused with the primary pitch [51]. A real-time psychoacoustic model evaluates the perceptual accessibility of each overtone by accounting for factors such as perceived loudness, resolvability, and masking. The system then isolates, amplifies, and plays back the most perceptually accessible overtone to the user.

We conducted two studies to evaluate the proposed system. In Study #1 (n = 32), we adopted a psychoacoustic approach to determine whether the overtone textures identified by our system could be effectively cued. We utilized a two-interval, two-alternative forced-choice (2I-2AFC) paradigm to quantitatively assess how a single, one-off cueing could influence perceptual experiences. Aggregated over 10 everyday sounds, participants showed a significant increase in detection of overtone texture immediately after cueing compared to the baseline, with accuracy improved by 13%. In contrast, a control group that received a sham cue (a tone shifted from the target overtone) showed no such improvement, confirming that the effect is not due to general task familiarization. The improvement in detection was no longer observed 5 minutes after, suggesting a short-term effect induced by brief cueing.

In Study #2 (n = 12), we conducted a within-subjects study to evaluate the device use in-lab. Participants interacted with three common household objects (e.g., an electric toothbrush) before, during, and after the system amplified an overtone texture. In 25 of the 36 trials (69%), participants reported that the cued texture became more perceptually salient after the device was turned off, as compared to the pre-device session. The attentional shift was evidenced by their drawings and descriptions of the sounds. Through analyzing participants' self-report responses, we also observed that reperceiving everyday sounds with hidden details has the potential to prompt new mental associations with familiar sounds.

We envision that hearables scaffolding users' attention toward overlooked acoustic textures beyond device use could enable several applications. First, it could support users in leveraging finer textures in everyday sounds as anchors for mindfulness in daily activities. In stressful acoustic environments, such as MRI scanners, such systems could preemptively cue users to direct their attention toward less overwhelming acoustic textures. Finally, such systems could help everyday listeners cultivate a deeper appreciation of music by drawing attention to subtle timbre details that are often only available to trained ears.

2 Related Work

2.1 Hearables that Mediate Everyday Perception

Hearables can be broadly categorized into two types: (1) assistive devices that compensate for hearing loss, such as hearing aids and cochlear implants [11, 12, 45, 72], and (2) devices that alter or augment the auditory environment for task-based purposes, such as media consumption [25, 54, 69]. This paper focuses specifically on the second category of consumer hearables.

Devices such as earbuds and headphones have become increasingly ubiquitous in everyday life [25]. These systems actively mediate how we perceive the real world [59]. However, just as smartphones draw users' attention toward digital content and reduce awareness of physical surroundings, the mediation of auditory perception raises important concerns in digital well-being.

When digital auditory content, such as music and podcasts, competes with environmental sounds for attention, hearables often prioritize immersion by effectively "canceling out" real-world sounds [28, 54, 69]. Such media immersion can reduce users' awareness of environmental changes, thereby increasing risks during activities such as walking in traffic or crossing roads [31, 34, 41].

Even in "transparency" mode—where external sounds are captured, processed, and played back to the user in real time without active noise cancellation—perception of real-world auditory cues remains compromised [6, 48, 66]. For example, Watanabe et al. reported that sound localization accuracy decreases from 91.5% with unaided ear to 68.9% when using hearables on transparency mode [66].

2.2 Benefits of Noticing Auditory Details

While hearable devices often cancel out real-world sounds, the ability to notice rich sensory cues in our environment can offer many benefits. One direct association is with mindfulness practices, cultivating attention and openness to present-moment experiences [7, 10]. Being able to notice present-moment auditory cues, from one's own breath to the movement of a hand, can help redirect users' thoughts from ruminating about the past or the future back to the present moment, thus creating a sense of calmness [15, 20, 32, 73]. Mindfulness practices are found effective to intervene emotional distress [46, 56] and enhance psychological well-being [37, 53]. Similarly, directing attention to natural soundscapes, such as birdsong, has been shown to facilitate recovery from stress [14, 55]. More broadly, attending to the sensory richness of everyday experiences has been identified as a fundamental component of beautiful experiences, with multisensory engagement that captures and sustains attention playing a central role [27, 35].

From an auditory perspective, the perception of sound qualities, such as timbre and pitch, also constitutes a form of musical listening[16, 51]. In this context, everyday sounds can become a source of musical appreciation [5]; musicians frequently draw inspiration from these acoustic components, incorporating them into compositions and improvisation practices [67, 70].

2.3 Hearables to Deepen Sensory Awareness

How can hearables deepen users' awareness of rich auditory cues in everyday sounds, rather than suppressing them?

Prior work tackling this question has focused on how hearables can augment user awareness *during* the device use. This includes interfaces that help users perceive speech in noisy environments [8, 9, 65]. For example, SoundBubbles demonstrates how spatial filtering can isolate and enhance specific speakers [9]. Emerging research has further explored heightening awareness of non-speech soundscapes. For instance, Semantic Hearing [64] enables users to selectively filter environmental sounds using text input, such as suppressing traffic noise to foreground bird chirping. Tao et al. developed a wearable system that amplifies subtle sounds produced by hand movement (e.g., fingers sliding through surfaces) to encourage attention to ongoing actions [60]. GutIO transforms abdominal sounds into haptic feedback to enhance interoceptive awareness of internal bodily sensations [42].

While prior work has shown great promise, the heightened awareness enabled by these systems remains only effective while the device is in use: once the hearable is removed, the rich auditory cues it reveals become inaccessible again. We instead explore an alternative role for hearables as temporary scaffolds that guide users to notice the rich acoustic textures of everyday sounds even *after* the device is removed. As an initial step toward this broader goal, we focus on short-term attentional shift as users hear the same sounds (i.e., reperceive) after device use.

3 Working Principles: Cueing to Reveal Hidden Texture in Everyday Sounds

Our work builds upon prior literature in auditory cueing to explore how hearables can scaffold users to notice rich texture in everyday sounds beyond device use.

In psychoacoustics, a cue acts as a guiding stimulus that directs listeners' attention toward an appropriate frequency, thereby reducing uncertainty when processing auditory information [18, 49]. This mechanism is frequently utilized to investigate the human capacity to detect, or "hear out," target tones embedded in a mixture of noises [23, 26, 39, 40, 47]. In these previous studies, cueing effects are typically evaluated in highly controlled setups using synthetic stimuli, such as sinusoidal tones embedded within carefully constructed masking noise, to understand the perceptual limits. Researchers have explored a range of cueing mechanisms, including visual cues, multitone signals or cues based on tonal patterns [23, 24, 26]. Among these, providing an auditory preview that matches the exact frequency of a target component has been shown to be particularly effective in facilitating its detection [26, 49]. Moreover, prior work has examined how different pulsing structures—such as presenting the cue or target as a rhythmic sequence rather than a continuous sound—further influence the detectability of these information [39, 40].

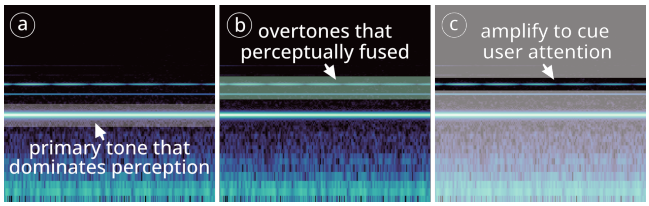


Figure 2: Spectrogram of the from a singing bowl hit. The x-axis represents time, and the y-axis represents frequency.(a) Normally, our perception is dominated by the fundamental tone. (b) The overtones that lie above it are perceptually fused. (c) Through cueing, our system amplifies the most perceptually accessible overtone.

We extend cueing techniques to everyday listening contexts with hearables. As the first step, we focus on the overtone texture in everyday sounds. Figure 2(a) shows the spectrogram of the sound produced by a singing bowl hit, a common household object used for mindfulness practices. The acoustic profile is dominated by its fundamental frequency at approximately 546 Hz. However, the singing bowl also creates a series of higher-frequency overtones that lie above this fundamental, as shown in Figure 2(b). For instance, the second-order overtone at around 1540 Hz exhibits a distinct pulsing structure. For everyday listeners, these overtones are perceptually fused with the main pitch and remain effectively "hidden."

Through cueing, our device identifies the most perceptually accessible overtones (in the singing bowl example, the 2nd order overtone) and guides users' attention to these previously hidden acoustic textures. Specifically, the system isolates individual overtones and amplifies them by applying digital gain. The initial prototype identifies the overtones that are the most perceptually accessible by considering loudness, resolvability and masking. By playing back these texture to users in real time, the device provides a cue

analogous to auditory preview techniques in the psychoacoustics literature. The goal is to induce a short-term attentional shift. After the device is turned off and users listen to the same signing bowl sounds again (i.e., reperceive), they now hear the higher-frequency overtones becoming more salient.

4 System Implementation

4.1 Hardware Setup

Figure 3 illustrates the hardware setup of the system. The input device consists of a neckband equipped with two microphones; specifically, we use Sonic Presence SP15C microphones [4] to capture stereo audio. The stereo audio preserves the spatial characteristics of everyday sounds, ensuring realistic listening experience for the users. For output, we use open-ear headphones to play back the selected texture to users. This design allows users to perceive enhanced texture layers while maintaining access to the natural acoustic environment.

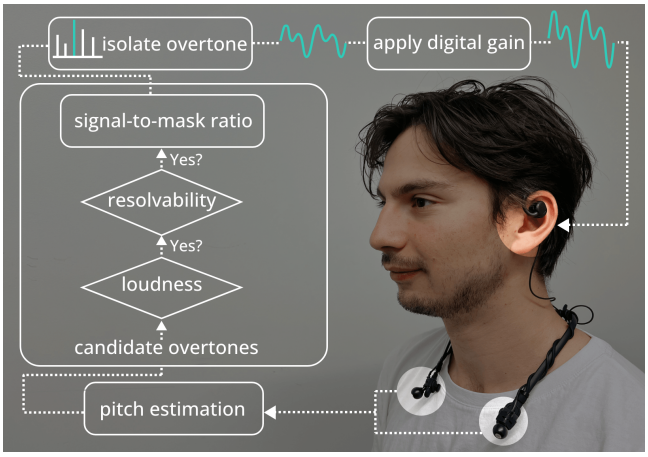


Figure 3: System Diagram. The system detects ongoing sounds using a neckband with stereo microphones. It then performs pitch detection to search for candidate overtone textures to be cued. The system evaluates the perceptual accessibility of each candidate overtone by considering loudness, resolvability, and masking. The highest-ranked overtone is then isolated and digitally amplified for playback to the user.

4.2 Pitch Estimation

To identify hidden overtone texture in sounds, our system first estimates the tone that dominates subjective qualities, also commonly referred to as pitch [21]. In most cases, pitch can be quantified by the fundamental frequency (f_0), which is the lowest—and often the most perceptually prominent—frequency in a harmonic series. Specifically, we utilize `SwiftF0` [43], a deep learning–based pitch estimation model, to generate initial f_0 candidates. The audio input is sampled at 44100 Hz and segmented into sliding analysis windows of $N = 4096$ samples with a hop size of 1024 samples.

Because everyday acoustic environments are highly dynamic, we implement a two-stage filtering pipeline to ensure stability of the output. We first discard estimates with a confidence score below 0.8, which decides based on heuristic tuning. We then apply a median filter of size 5 to reject transient outliers. The resulting estimates are then smoothed using a scalar Kalman filter, with the process noise covariance set to $Q = 0.1$ and the measurement noise covariance set to $R = 5.0$. This reduces jitter while maintaining responsiveness to changes in pitch.

4.3 Identify Candidate Overtones

Once the pitch is established, the system identifies candidate overtones in the sounds. Tones that are integer multiples of the fundamental frequency are most susceptible to perceptual fusion, as the auditory system integrates them into a single pattern [38]. Natural sounds often exhibit inharmonicity, meaning that overtones do not align perfectly with integer multiples of the fundamental. To account for this, our system performs a spectral search in regions centered around integer multiples of the fundamental frequency to identify actual energy peaks in the signal.

Specifically, given the estimated pitch f_0, for each overtone rank n, we define a search region: $f_{search} \in [n \cdot f_0 \pm 0.4 \cdot f_0]$. This search range is chosen to accommodate the inharmonic nature of everyday sounds while ensuring that adjacent search windows do not overlap.

4.4 Perceptual Accessibility of Overtones

Given the candidate overtones f_n, our system evaluates its likelihood of being perceptually accessible after cueing, by considering three psychoacoustic factors.

4.4.1 ***Loudness.*** An overtone cannot be perceived if it lacks sufficient energy. Therefore, the first factor we consider is the perceptual loudness of candidate overtones. To assess this, we compute the perceived loudness within the overtone's spectral bandwidth, adjusted for human frequency sensitivity using A-weighting.

Given a detected overtone f_n, how it is perceived is determined by the ear's perceptual bandwidth. We utilize Equivalent Rectangular Bandwidth (ERB) model [17] to approximate the auditory filter width at frequency f_n. The physical energy of an overtone is integrated within its corresponding ERB bandwidth. Let R_{ERB} denote half of this bandwidth ($R_{ERB} = ERB(f_n)/2$), and let $X[f]$ denote the frequency spectrum of the current audio frame. The physical power level on the dB scale is calculated by summing the power spectrum $|X[f]|^2$ across this range:

$$L(f_n) = 10 \cdot \log_{10}\left(\sum_{f=f_n-R_{ERB}}^{f_n+R_{ERB}} |X[f]|^2\right) \quad (1)$$

To approximate the human ear's diminished sensitivity to lower frequencies, we apply a standard A-weighting filter $W_A(f_n)$ [3]. The final perceived loudness is the sum of the integrated physical energy and this frequency-dependent weighting:

$$L_{perceived}(f_n) = L(f_n) + W_A(f_n) \quad (2)$$

The overtone is considered *audible* only if this level exceeds a calibrated noise floor threshold (set to −40 dBFS for our system hardware).

4.4.2 ***Resolvability.*** The human ear cannot distinguish between two frequencies if they are too close together such that they fall within the same critical band. Thus, the second psychoacoustics factor we consider is the spacing between overtones. An overtone

is considered "resolved" only if the spectral spacing between overtones is wider than the ear's critical bandwidth at that frequency. Unresolved overtones becomes perceptually impossible to be heard as distinct tones.

$$\mathbb{I}_{resolved} = \begin{cases} 1 & \text{if } f_0 > ERB(f_n) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

4.4.3 ***Masking.*** Even if an overtone region is perceptually loud and spectrally resolved, its accessibility is also influenced by the presence of other overtones or sounds at the same time. This effect, known as simultaneous masking, represents the third psychoacoustic factor we consider when evaluating overtone perception.

We estimate simultaneous masking using a simplified critical-band masking model based on classical psychoacoustic spreading functions. Advanced critical-band masking approximations are used in perceptual audio-coding models for data compression, such as MPEG-1 Psychoacoustic Model 1 [44, 74]. We first calculate the energy distribution across the audio spectrum. Following prior work to streamline computation for real-time processing, we group the Fourier transform bins into $N_B = 24$ critical bands on a Bark Scale [75]. We then approximate masking spread using asymmetric spreading slopes, with a fixed lower slope of 27 dB/Bark and a level-dependent upper slope, following prior work [61, 62, 76]. The masking level $M_{i,j}$ imposed by band j onto band i is calculated by subtracting the Bark-distance-dependent spreading loss and fixed masker-to-threshold offset $O_{mask} = 10dB$ from the masker band level $L_{band}(j)$. Since multiple maskers contribute to each band simultaneously, the individual masking effects are additive in the power domain. The global masking threshold $T(i)$ for the i-th band is calculated by summing the energy of all spreading functions acting upon it:

$$T(i) = 10 \log_{10} \left(\sum_{j=1}^{N_B} 10^{\frac{M_{i,j}}{10}} \right) \tag{4}$$

An overtone with a level below this global threshold is considered masked. To quantify it, we calculate the Signal-to-Mask Ratio (SMR) for the n-th overtone by comparing its power level $L(f_n)$ against the global threshold at its corresponding frequency band i_n:

$$SMR_n = L(f_n) - T(i_n) \tag{5}$$

A negative SMR indicates that the overtone level falls below the estimated masking threshold and may therefore be perceptually masked. Because the masking estimate is approximate and detectability near the threshold is uncertain, we use a decision margin of 3dB.

4.4.4 ***Selection Logic.*** The final perceptual accessibility score S_n synthesizes these three psychoacoustic factors—loudness, resolvability, and masking—into a single ranking metric to determine which overtones the system should amplify.

First, if an overtone fails either the loudness check ($\mathbb{I}_{\text{audible}} = 0$) or the spectral resolvability check ($\mathbb{I}_{\text{resolved}} = 0$), it receives a severe penalty of −100. This effectively eliminates it from consideration, as the ear cannot physically perceive it regardless of cueing.

For overtones that pass the loudness and resolvability checks, their baseline score is dictated by their Signal-to-Mask Ratio (SMR_n). If an overtone is masked by adjacent frequencies ($\text{SMR}_n < -3$ dB), the system applies a moderate penalty of −50 to lower its ranking. Finally, overtones that are both highly resolved and sufficiently prominent against the masking threshold ($\text{SMR}_n \geq -3$ dB) retain their raw SMR_n as their final score. This ensures the system prioritizes amplifying the most perceptually accessible overtones in the user's environment.

4.5 Cueing Users with Hidden Texture

Once the target overtone is selected, the system isolates the highest-ranked overtone using a band-pass filter. The filter is centered at the detected peak frequency f_n with a fractional bandwidth of 0.15. The selected frequency region is then amplified by applying a digital gain of user's choices.

4.6 Latency Measurement

In a technical evaluation, we measured the system's end-to-end latency. The total latency primarily arises from two sources: (1) pitch detection used to estimate candidate overtones, and (2) audio streaming latency during playback of the amplified overtones.

The pitch estimation model, *SwiftF0*, reports a runtime of approximately 132 ms on a standard desktop computer with a modern CPU [43]. While pitch detection latency affects how quickly the system can identify the most accessible overtones, streaming latency directly impacts the listening experience.

To isolate streaming latency, we configured the system to lock onto the 2nd overtone. We played a composite tone at a distance of 10 cm from the system's microphone. The output was recorded using two methods: (1) capturing the sound played back through the open-ear headphones after it passed through the full processing pipeline, and (2) recording the same ringtone using a separate microphone placed near the audio source as a baseline. Latency was computed by comparing the temporal alignment of the two recordings and averaging the results across 30 trials.

We found that the streaming latency is 24.6 ms (SD = 1.69). This value falls below the just noticeable difference (JND) threshold for audio latency in musical applications, which is typically 30 ms and above [30, 36, 50]. While some studies have measured optimal minimal perceivable thresholds as low as 10 ms [68], these thresholds are often determined in highly controlled experimental settings. Such conditions do not necessarily reflect human performance in real-world situations, where attention is divided among multiple tasks and users need to navigate a multitude of sensory signals.

5 Study #1: Psychoacoustic Study

Can the overtone textures identified by our system be effectively cued? To answer this question, we conducted an in-lab study using psychoacoustic methods to quantitatively measure changes in participants' perceptual experiences.

5.1 Study Design

5.1.1 ***Two-interval Two-alternative Forced Choice Paradigm.*** To systematically measure whether certain auditory details could be perceived by participants, we adopted a two-interval, two-alternative forced-choice (2I-2AFC) paradigm, following prior work [57]. In each trial, participants were presented with two successive sound intervals: the original sound and a modified version of the same sound with an overtone component removed, as shown in Figure 4.

Participants were asked to indicate which interval contained a tone that the other did not. An inability to discriminate between the two sounds (i.e., accuracy at chance level) indicates that the removed overtone was perceptually "hidden" to the listener. In each trial, the target overtone was the component identified by our system as being the most perceptually accessible within the sounds, with ranking details outlined in Section 3.

After each trial, participants rated the perceived mental effort required for the task and their confidence in their answer on a 1 (low) to 5 (high) Likert scale [13].

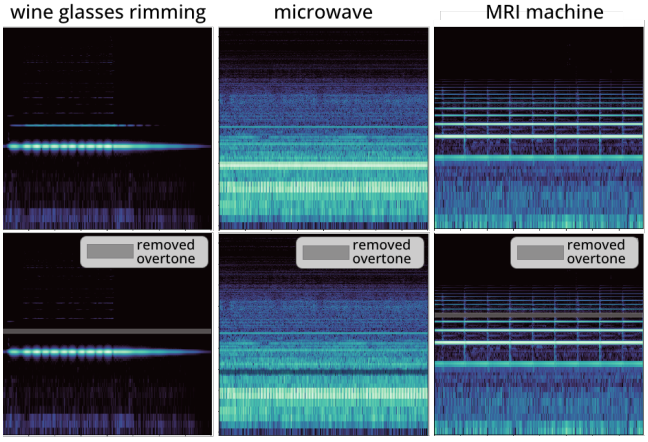


Figure 4: Demonstrating the comparison stimuli pair used in Study #1. Top row shows the original sounds spectrogram and the bottom row shows the spectrogram with the target overtone removed. The x-axis represents time, and the y-axis represents frequency.

5.1.2 ***Repeated Measures.*** The aforementioned comparison task was repeated in three different sessions. To examine whether participants can naturally hear out the target overtone, they first completed the task before any auditory cueing (*baseline*). To evaluate the immediate effect of auditory cueing on overtone perception, participants then repeated the same task immediately after hearing the cueing tone (*post-cue*). To probe whether the cueing effect persist, participants completed the task again five minutes after the *post-cue* session (*5-min-post-cue*).

5.1.3 ***Between-Group Design.*** Due to the repeated-measures design of the study, we accounted for potential effects of procedural learning by including a sham control group. Participants in this group were informed that they would hear a tone intended to assist with the task, same as the experimental group. However, the provided cueing tone did not correspond to the overtone component present in the original sound. Instead, it was derived by shifting the target overtone used in the experimental condition by a multiple of 1.5 (i.e., a perfect fifth). This manipulation approximate the scenario in which participants are cued to attend to a pitch that is not present or not perceptually accessible in the stimulus.

5.1.4 ***Stimuli.*** To increase the generalizability of the results, we sampled 10 everyday sounds with perceptually meaningful overtone structures. These stimuli included: sounds from an electric toothbrush, hairdryer, HVAC fan, low bell, microwave, mosquito, MRI machine, refrigerator, singing bowl, and wine glasses rimming. Together, these stimuli span a broad range of sound-producing objects, from musical instruments to household appliances and clinical machines. All audio stimuli are from online database FreeSound [2] and have at least one overtone. To avoid perceived loudness change after overtone removal, all sounds were loudness-normalized to -25dBFS. All audio clips are set to be 8 seconds long.

Before the main experiment, all participants completed five practice trials to familiarize themselves with the task. Each practice trial presented a pair of stimuli, one consisting of a single pure tone and the other consisting of two pure tones. Participants were asked to identify which stimulus contained an additional tone that was absent from the other stimulus, following the same procedure as in the main experiment. Participants were required to correctly complete all five practice trials before proceeding to the experimental sessions. All stimuli used in the experiment are provided in the Supplementary Materials.

5.1.5 ***Randomization & Counterbalancing.*** Participants were randomly assigned to either the control or experimental group. Each comparison pair was presented twice within each session block (baseline, post-cue, and 5-min-post-cue), with the order of the intervals counterbalanced. Within each session block, the presentation order of the comparison pairs was randomized. In each session, participants completed 20 trials of comparisons and 60 trials in total for the whole study.

5.2 Hypotheses and Research Questions

We formulate two hypotheses. Our first hypothesis concerns the detection performance within the experimental group. We hypothesize that the overtone predicted by our system is more likely to be detected immediately after cueing (*post-cue*) compared to the baseline session (H1). Between the two condition group, we also hypothesize that immediately after the cueing (*post-cue*), participants in the experimental group will perform better in the detection task than in the control group (H2).

As a research question, we investigate the persistence of the cueing effect (RQ1): Does the cueing effect persist five minutes after? In addition, we explore how exposure to the overtone identified by our system influences cognitive load and confidence in detecting hidden texture. Specifically, we ask: How does exposure to the detected overtone affect participants' cognitive load during the task (RQ2)? How does exposure to the detected overtone affect participants' confidence in their judgments during the task (RQ3)?

5.3 Participants

We recruited 33 participants who were over 18 years old and reported no hearing deficits and no formal ear training. One participant's data was excluded due to incomplete data, resulting in 32 participants (16 per group) for analysis. Specifically, 21 self-identified as females, 11 self-identified as males. The average age of participants was 24.2 years old (SD = 4.75). 4 participants identified as Caucasian, 3 identified as African American, 1 identified as Native American, 22 identified as Asian, 2 chose "other". Note that participants could select more than one ethnicity, resulting the total number here exceeding the sample size.

Table 1: detection accuracy (%) per sound type, by session and condition.

	bell tower	hairdryer	HVAC fan	toothbrush	microwave	MRI	singing bowl	wine glasses	mosquito	refrigerator
baseline (control)	59.4%	56.2%	78.1%	50.0%	28.1%	59.4%	62.5%	40.6%	59.4%	56.2%
post-cue (control)	53.1%	53.1%	50.0%	59.4%	53.1%	56.2%	56.2%	62.5%	46.9%	53.1%
5-min-post-cue (control)	53.1%	71.9%	65.6%	43.8%	50.0%	50.0%	56.2%	62.5%	59.4%	53.1%
baseline (experimental)	53.1%	43.8%	71.9%	43.8%	28.1%	50.0%	43.8%	53.1%	65.6%	43.8%
post-cue (experimental)	65.6%	68.8%	84.4%	84.4%	59.4%	56.2%	50.0%	59.4%	59.4%	43.8%
5-min-post-cue (experimental)	59.4%	59.4%	71.9%	53.1%	53.1%	50.0%	56.2%	62.5%	56.2%	40.6%

The study took approximately 60 minutes and was approved by Stanford Institutional Review Board (IRB # 71992). Each participant was compensated $30 USD for their time.

5.4 Results

Due to the binary choice nature of the task, we analyzed the detection accuracy using a Generalized Linear Mixed Model with a binomial distribution and logit link function. We factored in condition group (*experimental*, *control*), session (*baseline*, *post-cue*, *5-min-post-cue*), and their interactions as fixed effects. Participant identifier and audio stimuli identifier were included as random effects to account for individual differences and variability across sounds. For the confidence and mental effort ratings, we used a linear mixed-effects model with the same factor configurations. Table 1 and Figure 5 showed the results of Study #1.

5.4.1 ***Manipulation Check of the Baseline Detection Accuracy.*** We first examined baseline accuracy in both groups, which shows no significant difference ($\beta = -0.22$, $p = .215$). *Control* group had a baseline accuracy of 55.0%, while *experimental* group achieved 49.7% accuracy. The results indicated that participants in both group performed similarly at a chance level on average and the target overtone was initially "hidden".

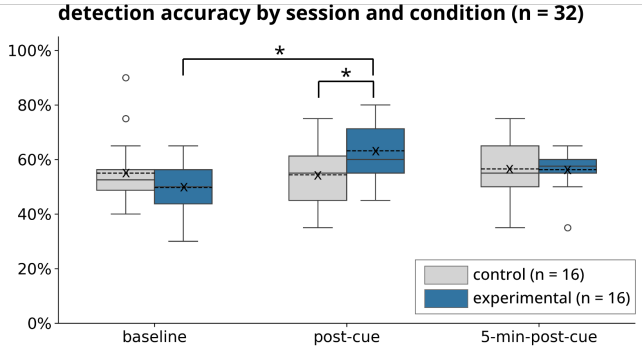


Figure 5: Study #1 results on detection accuracy.

5.4.2 ***Cueing Improves the Ability to Detect Overtones in Everyday Sounds.*** A significant interaction was found between group condition and session. Specifically, the *experimental* group showed a significantly different trend during the post-cue session compared to the *control* group ($\beta = 0.59$, SE = 0.23, z = 2.61, $p = .009$). Post-hoc pairwise comparisons using Bonferroni correction were conducted to further explore these differences both between the two conditions and within the experimental group.

Within the *experimental* group, we found that participants performed significantly better during the *post-cue* session compared to the *baseline* session ($\beta = -0.56$, $SE = 0.16$, $z = -3.45$, $p = .002$). Accuracy increased from 49.7% at baseline to 63.1% during the cueing session, indicating that participants were more likely to perceive the previously hidden overtone after cueing. The effect size was calculated with the odds ratio [58]. We found that the experimental group had 1.75 times higher odds of detecting the overtone texture immediately after cueing compared to baseline (OR = 1.75), corresponding to a moderate increase in detection with a single cueing. Specifically, detection accuracy increased in *post-cue* for 8 of 10 sounds relative to baseline for the *experimental* group. These results support H1, which shows that participants are more likely to hear out these hidden texture immediately after the cueing.

Compared to the *control* group, which received a sham cueing tone (with 54.4% average accuracy), participants in the *experimental* group performed significantly better at identifying the hidden overtone ($\beta = -0.37$, $SE = 0.18$, $z = -2.06$, $p = .039$), with odds ratio of 1.45. In contrast to the *experimental* group, only 3 of the 10 sounds showed improved detection accuracy in the *control* group. This result supports *H2*, indicating that the improvement was specific to the targeted cue rather than a general effect of procedural learning.

5.4.3 ***No Persistence of Cueing Effect Found 5 Minutes After.*** No significant interaction was found between condition group and the *5-min-post-cue* session ($\beta = 0.21$, $SE = 0.23$, $z = 0.92$, $p = .358$). This suggests that the brief cueing duration (8 seconds) leads to change in perceptual experiences immediately after but was insufficient to produce a longer-lasting perceptual change.

5.4.4 ***The Success of Cueing Came with Higher Mental Effort.*** At *baseline*, there was no significant difference in mental demand between the groups ($t(31.15) = -1.30$, $p = .20$). The model revealed a significant interaction between group condition and session, at both the *post-cue* ($t(1875) = 2.77$, $p = .006$) and *5-min-post-cue* ($t(1875) = 2.81$, $p = .005$) sessions. Post-hoc comparisons utilizing Bonferroni adjustment confirmed that within the *experimental* group, participants' mental demand was significantly higher in the *post-cue* session ($M = 2.83, SD = 1.05$) compared to their baseline ($M = 2.49, SD = 0.68; p < .001$). Similarly, a significant increase in mental demand was observed in the *5-min-post-cue* session ($M = 2.77, SD = 1.07$) compared to the *baseline* (p < .001). Conversely, no significant differences were found within the control group across the *post-cue* ($M = 3.05, SD = 1.04$), *5-min-post-cue* ($M = 2.99, SD = 1.04$), and baseline sessions. No significant differences was found between two groups at *post-cue* ($p = .52$) and *5-min-post-cue* ($p = .53$) sessions on the reported mental demand.

5.4.5 ***Cueing, Both Real and Sham, Decreased Confidence.*** At *baseline*, there was no significant difference in reported confidence between the *experimental* and *control* groups ($t(33.65) = 1.19, p = .242$). The model revealed a significant main effect of session at the first post-cue interval (*post-cue*), where participants' confidence scores decreased significantly compared to *baseline* ($t(1875) = -2.31, p = .021$). The lack of a significant interaction at this session ($p = .60$) indicates that this initial decline in confidence was consistent across both groups. Specifically, participants in the control group reported a mean confidence rating of 2.50 ($SD = 0.70$) at *baseline*, 2.33 ($SD = 0.86$) at *post-cue*, and 2.56 (SD = 0.87) at *5-minute post-cue*. Participants in the *experimental* group reported a mean confidence rating of 2.82 ($SD = 0.69$) at *baseline*, 2.58 (SD = 0.69) at *post-cue*, and 2.67 (SD = 0.70) at *5-minute post-cue*.

5.4.6 ***Summary of Study #1 Findings.*** We showed that hidden texture identified by our system can be successfully cued. When participants received a cue that previewed the overtone, their detection accuracy significantly improved compared to baseline performance. In contrast, the *control* group that received a sham cue did not show the same level of improvement, indicating that the effect cannot be attributed to procedural learning. However, successful cueing was also associated with increased perceived mental effort. This increase may reflect greater engagement with the detection task where guiding user attention to frequencies they would not otherwise pay attention to but with meaningful information. On the confidence level, the decrease of it in both groups could have been due to initially overestimating their ability to detect until exposing to the reference cues.

6 Study #2: System Study in Everyday Listening

Building on the findings from Study #1 that leverages a psychoacoustics approach, Study #2 extends the evaluation by examining how the system influences participants' perception of the same sounds after using the device in a controlled, in-lab setting. In the study, participants interacted with everyday objects that produce sounds with rich overtone textures.

6.1 Study Design

We adopted a within-subjects pre–post study design in which all participants followed the same protocol. Participants were invited to the lab and interacted with three objects in a counterbalanced order: an electric toothbrush, a singing bowl, and a vibrating phone.

For each object, participants interacted with it and listened to its sounds for three sessions: (1) a 1-minute *pre-device* session with the device turned off; (2) a 1-minute *device-cueing* session with the system turned on and cueing, during which the system amplified the predicted overtone; and (3) a final 1-minute *post-device* session with the device turned off again.



Figure 6: Objects used in the Study #2. (a) a vibrating phone, (b) an electronic toothbrush and (c) a singing bowl.

6.2 Apparatus

Figure 6 depicts the objects participants interacted with. All objects were selected for their latent overtone structures, exhibiting multiple overtones above the fundamental pitch.

Electric toothbrush. We used a commercially available household device, the *7MAGIC Sonic Electric Toothbrush*. The toothbrush was set to the "sensitive" mode at its highest intensity. In this mode, the toothbrush produces a periodic up-and-down motion, generating a low-frequency fundamental pulsing at approximately 310 Hz, along with higher-frequency harmonic overtones. The device detected second-order overtone at 926 Hz as the most accessible.

Vibrating phone. An iPhone 13 Pro was placed on top of a cardboard box. The phone vibrated in a consistent pattern at approximately 150 Hz, corresponding to the dominant frequency of the emitted sound. In addition to this fundamental component, the vibration produced several higher-order overtones and the 4th overtone at 754 Hz is the one selected by the system to be amplified.

Singing bowl. When struck on the side, the singing bowl produces a consistent low-frequency tone at approximately 546 Hz, accompanied by several higher-order overtones. The device detected second-order overtone at 1540 Hz as the most accessible overtones and amplifies it. During the interaction, participants were instructed to strike the bowl using a wooden mallet in a consistent manner. After each strike, participants waited for the sound to decay before initiating the next interaction.

Participants were instructed to interact with the objects consistently to ensure comparable sound production across sessions. The amplified overtone was kept the same for all participants to provide a consistent experience. Its amplification level was set to 24 dB, a value chosen to compensate for energy loss introduced by band-pass filtering and to make the overtone salient enough relative to the original sound.

6.3 Measures

6.3.1 ***Perception of Everyday Sounds.*** To understand how perceptual experiences of object sounds changed before and after using the device, participants were asked to both draw and describe the sounds they heard. In the drawing task, participants represented each perceived pitch as a separate line, with higher pitches placed higher on the page. Line thickness indicated perceived loudness. This representation method was adapted from prior work [29], which found that both musicians and non-musicians perceive such pitch and loudness mappings as intuitive. These drawings enabled researchers to visually examine changes in participants' perception of object sounds.

In addition to the drawings, participants completed a questionnaire after both the *pre-device* and *post-device* sessions to describe what they heard from the objects.

6.3.2 ***Mental Association.*** Beyond perceptual experience, we also examined how the system changes interpretations of everyday sounds by focusing on mental associations. Participants completed pre- and post-session questionnaires in which they described what they just heard from the objects make them think about. We then coded all reported entities (e.g., objects, places, people, events) referred in their responses. Changes from pre-device to post-device

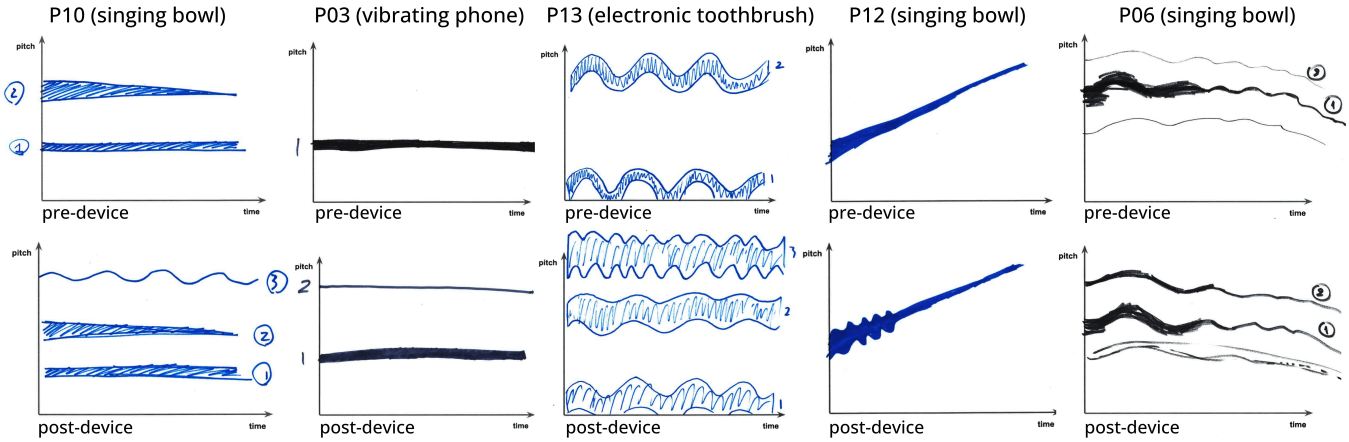


Figure 7: Representative drawings from participants in Study #2. The top row illustrates participants' perception of the object sounds in the *pre-device* session, and the bottom row shows their perception in the *post-device* session. Different lines represent distinct perceived pitches, and line thickness indicates the loudness of each tone.

associations were used to capture how participants' mental associations with the sounds shifted as a result of using the device.

6.4 Participants

We recruited 14 participants who were over 18 years old and reported no hearing deficits. Two participant's data was excluded due to incomplete trials, resulting in 12 participants for analysis. Specifically, 6 self-identified as females, 5 self-identified as males, and 1 self-identified as non-binary. The average age of participants was 25.3 years old (SD = 3). 1 participants identified as Caucasian, 9 identified as Asian, 1 chose "other" and 1 chose "prefer not to say".

The study took approximately 75 minutes and was approved by Stanford Institutional Review Board (IRB # 71992). Each participant was compensated $25 USD for their time.

6.5 Results

6.5.1 ***Finding #1: Our System Directs User Attention to Overlooked Details of Everyday Sounds.*** Consistent with the findings from Study #1, participants reported that the overtone cued by the device became more perceptually salient after the device was turned off, as compared to before use. This shift in attention was coded by jointly analyzing participants' sound drawings and written descriptions. A successful instance of cueing was defined as drawings and descriptions that demonstrated either (1) additional textures aligning with overtone the device cued, or (2) increased prominence of the cued overtone. Responses were not counted as successful cueing if participants simply repeated their initial descriptions or elaborated with additional details unrelated to the device-cued features.

Across all 36 trials (12 participants × 3 objects), participants reported the overtone texture cued by our device becoming more perceptually salient after device use in 25 instances, accounting for 69.4% of the trials. Eleven out of twelve participants perceived the hidden texture as more salient for at least one object. A breakdown by object revealed 9 out of 12 successful trials for both the vibrating phone and the singing bowl, and 7 out of 12 successful trials for the electric toothbrush.

Figure 7 presents participants' drawings visualizing the sounds they perceived from the objects before and after using the device. Changes in perceptual salience manifested in two main ways. In the majority of cases (18 out of 25), participants did not detect overtones during the pre-device session. Following device use, cueing was reflected in the emergence of additional texture in their drawings, as illustrated in columns 1–3 of Figure 7.

For instance, P01 described the vibrating phone sounds in the *pre-device* session as *"I hear the buzzing of the phone. The sound is quite stable (single pitch), the oscillations in the sound is so minimal I barely hear it. It also sounds like this is not just a single pitch, but the rest is so faint I cannot identify."* After using the device, when P01 listened to the same sounds, they reported, *"I hear the very slight overtone (of higher pitch). This is a sound I did not hear in the first session, but I only hear after listening to the sound with the device."* Similarly, for the singing bowl, P10 shared: *"...the higher pitched sound of the singing bowl was noticeably higher than it was in the first session. I think my ears were more sensitive to the higher pitched sound after using the auditory device."*

In this context, the increased perceptual salience of the overtone could also contribute to the sounds to be more holistically "textured". Figure 7 (column 4) illustrates one such example: P12 initially described the singing bowl sound as transitioning smoothly from a low pitch to a high pitch. After being cued by the device to attend to a pulsing overtone, they perceived the sound as exhibiting a pulsing quality throughout the transition. Similarly, for the phone vibration sounds, participant P08 continued to report perceiving a single pitch; however, they described the sound as becoming generally higher in pitch and more prominent after the device. These results align with prior literature suggesting that everyday listeners without musical training tend to describe sounds holistically rather than understanding them as individual pitches [51].

In 7 out of 25 cases (4 with the singing bowl), the cued overtones were already noticed during the initial session. In all of these cases, after using the device, the cued overtone textures became more prominent, with one drawing example showed in Figure 7 (column 5). This increased salience was reflected in how the overtone texture were depicted with greater prominence (i.e., thicker markings)."*This*

time I heard the higher pitch tone persist for longer than I had noticed before. I also thought it sounded a bit louder." as written by P03 in describing the singing bowl sounds after the device use.

6.5.2 ***Finding #2: Hearing More from Everyday Sounds Has the Potential to Spark New Mental Associations.*** How does being able to perceive everyday sounds with hidden texture change participants' associations with them? We assessed this by analyzing the entities participants reported that they associated with the objects sounds in the questionnaire. Specifically, we identified trials in which participants introduced new descriptive entities in the *post-device* responses that were absent from their *pre-device* responses.

In 18 of the 25 trials (72%) where participants perceived the overtone texture as more salient, they also reported new entities associated with the sounds. For instance, during the *pre-device* session interacting with a electronic toothbrush, P02 wrote: *"The sounds of an electric toothbrush make me think of either my morning or evening routine, since I use a similar electric toothbrush at home."* After cued to hear the pulsing overtones from the electric toothbrush, P02 then described the sounds as: *"I did not associate it with the routines as in the first session. Because of the slow pulsing, it reminded me of a massage chair and was kind of relaxing."* Similarly, P12, at first associated the singing bowl sounds as *"sound being a wave, like in a calm pond, traveling long distances."* After the device use, P12 wrote *"It now makes me think of a pendulum that goes back and forth, then finally dies down."* For the phone vibration, while the sound normally make participants think about notification or phone call, just as P05 described in the pre-device session *"it brings people with a little bit nervous feelings because we often set up the clock with similar sounds to wake up, or like some urgent phone calls."* After the device use and notice the hidden overtone, P05 described the sounds as *"the vibrating phones sometimes sounds like the waves, I can [imagine] that it comes to the beach and goes back to the ocean, because it has kinds of vibrations."*

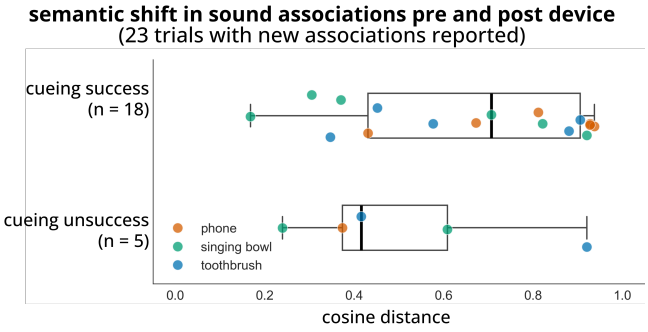


Figure 8: Semantic shift between entities associated with object sounds in *pre-device* and *post-device* sessions. This graph specifically focuses on trials where new entities emerged in the *post-device* session.

The introductions of new entities were also observed in 5 of 11 trials (45%) where device successfully cued participants to notice more texture. For example, participant P14 initially associated the phone vibration with a *"wake-up alarm"* and *"phone calls,"* and later added *"emergency fire alarm,"* despite no increase in overtone salience. Similarly, participant P07 expanded their associations for the singing bowl from *"meditation/chanting halls in a temple"* to include *"middle school meditation class"* and *"music session"* after the intervention.

To quantify how much participants' mental associations changed in trials where new entities were introduced, we computed the cosine distance between text embeddings of the reported entities in *pre-device* and *post-device* session, using the *all-mpnet-base-v2 model* [1]. The results showed an average distance of 0.65 (SD = 0.26) for trials in which the device cued successfully compared to 0.51 (SD = 0.26) for trials without successful cueing.

Although the sample size does not support formal statistical testing, these results suggest that device cueing could potentially encourage participants to form new associations with everyday sounds that are more semantically distinct from their initial interpretations.

7 Potential Applications

By leveraging auditory cueing to enable users to reperceive everyday sounds with hidden texture, our system opens up a range of potential applications, including mindfulness, reinterpreting stressful acoustic environments, and enhancing musical appreciation.

7.1 Everyday Sounds as Mindfulness Anchors

One potential application of our system lies in supporting users to leverage everyday sounds as anchors for mindfulness practices. Existing mobile technologies to support mindfulness often focus on structured, formal exercises, such as sitting meditation, that leverage a synthetic soundscape to guide user attention. This leaves everyday practices that weave mindfulness into daily activities under-supported [32, 33]. By actively revealing hidden textures in everyday sounds, our system lowers the barrier for users to enter into mindful states, staying attentive and curious about the present-moment experiences.



Figure 9: (a) A person is occupied by their phone while waiting for food to heat in the microwave. (b) By using our system to cue the user to hear out hidden acoustic textures within the microwave's hum, the idle time of waiting is transformed into a moment to foster curiosity and mindfulness.

Figure 9 illustrates one such example. A user is waiting for a microwave to heat up their food. Normally, the person is occupied on their phone, completely tuning out. With the device, the system isolates and amplifies the rhythmic, higher-frequency overtones hidden within the microwave's mundane hum. With a short seconds of exposure, the user also starts to notice these hidden textures they did not notice before. This temporary shift in attention draws them away from the screen, transforming an idle wait into a brief moment of mindfulness practice.

7.2 Reinterpreting Stressful Acoustic Stimuli

The ability to notice hidden textures could also offer users a new technique for navigating stressful acoustic environments by redirecting attention toward specific tonal components. For example, MRI scanners produce loud, repetitive sequences of pulsing that are often perceived as anxiety-inducing. By shifting attention from overall sound quality to specific overtones, users may form new associations with these sounds. For instance, higher overtones in the signal may be perceived as instrument-like rather than mechanical. Figure 10 illustrates this: a user wearing a VR headset undergoes training with a virtual MRI machine while also using our device, which detects and amplifies overtones to guide attention. This system can support preparatory training, enabling users to learn how to engage with their underlying acoustic structures in these stressful sounds before entering the actual scanner later without the device.



Figure 10: A user in preparatory training in a VR-simulated MRI environment. Our system detects and amplifies hidden overtones, teaching the user to redirect their attention toward these acoustic details to mitigate stress before an actual scan.

7.3 Musical Appreciation for Everyday Listeners

Through extensive training, musicians are able to "hear out" distinct acoustic components within complex compositions [63, 71]. In contrast, everyday listeners are often less able to appreciate music at this level of detail. Our system could serve as a short-term scaffold to support these listeners—who may lack formal training—in reperceiving music by guiding their attention toward subtle textures that typically blend into the overall mix. For example, within a multi-instrument passage, the device could guide users to reperceive music with the low hum of a bass guitar or the shimmering higher tones of a flute that might otherwise go unnoticed.

8 Limitations & Future Work

This paper serves as an initial step toward exploring how hearables can scaffold users to notice rich details in everyday sounds beyond device use, and it comes with several limitations.

On the system side, the current implementation serves as a proof-of-concept and uses a simplified psychoacoustic model to estimate hidden overtone textures in sounds. As a result, it primarily supports sounds with relatively stable and perceptually meaningful overtone structures in a controlled environment. To extend the system to more complex and dynamic environments, future work should consider additional perceptual factors, such as temporal masking and head-related transfer functions, to more comprehensively model human auditory perception. Moreover, many everyday sounds do not exhibit well-defined harmonic components, such as paper crumpling or footsteps. Similarly, while the proposed system focuses on overtones, they are not the only type of acoustic detail that may be perceptually fused and remain "hidden" in everyday sounds. Other acoustic features—such as resonance, reverberation, and percussive elements—are also present, yet they are not always attended to. Future work could expand the system to support a broader range of acoustic features, enabling users to perceive and engage with diverse forms of hidden auditory details.

Our psychoacoustic study demonstrated that the overtone textures identified by our system could be effectively cued, producing a short-term attentional effect. However, the current Study #1 only evaluates brief, one-time cueing and the long-term impact of repeated exposure remains unknown. Future work could explore training paradigms that provide feedback based on users' responses to investigate how and when perceptual learning emerges. Moreover, while our study evaluated 10 stimuli, this set remains relatively small compared to the diversity of everyday sounds. Even for the same type of sound source—for example, an electronic toothbrush—differences in hardware can produce variations. Similarly, the participant pool in this study was limited to young adults with no hearing deficit, and hearing capabilities can vary across age groups, which is not fully captured in the current evaluation. Future work should expand the analysis to include more diverse stimuli and a broader participant pool. Online crowdsourcing platforms, such as Prolific, could provide a promising venue for such investigations.

Study #2 extends our evaluation to device use while interacting with everyday objects in the lab. While the current study establishes the feasibility, the evaluation remains limited in ecological validity because it was conducted in a controlled lab setting. Future work should evaluate the system in more dynamic, real-world environments to examine whether the attentional effect persists. Moreover, in Study #2, while we uncovered initial effects suggesting that the device can encourage new associations with everyday sounds, the sample size is too small to support comparative conclusions. Future work should expand the evaluation to a larger participant pool.

9 Conclusion

In this work, we explore how hearables can help users discover hidden textures in everyday sounds. Grounded in auditory cueing, our device induces a short-term attentional shift that enables users to perceive these textures even after the device is removed. Across two user studies, we found that the system can effectively cue participants to detect previously overlooked overtone textures and the device use could potentially foster new mental associations with familiar sounds. We envision this work as a step toward hearables that not only mediate auditory experience, but also cultivate users' ability to attend to its richer details.

Acknowledgments

The authors would like to thank Libby Ye, Louise Lu, and Prof. Jeremy N. Bailenson for early conversations on this project. The lead author Yujie Tao is supported by the Stanford Knight-Hennessy Scholarship.

References

[1] 2026. all-mpnet-base-v2. https://huggingface.co/sentence-transformers/all-mpnet-base-v2.

[2] 2026. Freesound. https://freesound.org/.

[3] 2026. IEC 61672-1. https://webstore.iec.ch/en/publication/5708.

[4] 2026. Sonic Presence. https://www.sonicpresence.com.

[5] Jean-François Augoyard and Henri Torgue. 2006. *Sonic experience: a guide to everyday sounds*. McGill-Queen's Press-MQUP.

[6] Sebastian A Ausili, Nathan Erthal, Christopher Bennett, and Hillary A Snapp. 2025. Sound Localization with Hearables in Transparency Mode. *Audiology Research* 15, 3 (2025), 48.

[7] Scott R Bishop, Mark Lau, Shauna Shapiro, Linda Carlson, Nicole D Anderson, James Carmody, Zindel V Segal, Susan Abbey, Michael Speca, Drew Velting, et al. 2004. Mindfulness: A proposed operational definition. *Clinical psychology: Science and practice* 11, 3 (2004), 230.

[8] Ishan Chatterjee, Maruchi Kim, Vivek Jayaram, Shyamnath Gollakota, Ira Kemelmacher, Shwetak Patel, and Steven M Seitz. 2022. ClearBuds: wireless binaural earbuds for learning-based speech enhancement. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. 384–396.

[9] Tuochao Chen, Malek Itani, Sefik Emre Eskimez, Takuya Yoshioka, and Shyamnath Gollakota. 2024. Hearable devices with sound bubbles. *Nature Electronics* 7, 11 (2024), 1047–1058.

[10] J David Creswell. 2017. Mindfulness interventions. *Annual review of psychology* 68, 1 (2017), 491–516.

[11] Harvey Dillon. 2012. *Hearing aids*. Thieme Medical Publishers.

[12] Brent Edwards. 2007. The future of hearing aid technology. *Trends in amplification* 11, 1 (2007), 31–45.

[13] Danyang Fan, Walker Smith, Takako Fujioka, Chris Chafe, Sile O'Modhrain, Diana Deutsch, and Sean Follmer. 2025. Perceiving Slope and Acceleration: Evidence for Variable Tempo Sampling in Pitch-Based Sonification of Functions. *IEEE Transactions on Visualization and Computer Graphics* (2025).

[14] Luyao Fan and Mohamad Rizal Baharum. 2024. The effect of exposure to natural sounds on stress reduction: a systematic review and meta-analysis. *Stress* 27, 1 (2024), 2402519.

[15] Lisza Gaiswinkler and Human-F Unterrainer. 2016. The relationship between yoga involvement, mindfulness and psychological well-being. *Complementary therapies in medicine* 26 (2016), 123–127.

[16] William W Gaver. 1993. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological psychology* 5, 1 (1993), 1–29.

[17] Brian R Glasberg and Brian CJ Moore. 1990. Derivation of auditory filter shapes from notched-noise data. *Hearing research* 47, 1-2 (1990), 103–138.

[18] Ervin R Hafter, Robert S Schlauch, and Joyce Tang. 1993. Attending to auditory filters that were not stimulated directlya. *The Journal of the Acoustical Society of America* 94, 2 (1993), 743–747.

[19] Thich Nhat Hanh. 2016. *The Miracle of Mindfulness: Gift Edition*. Beacon Press.

[20] Adam W Hanley, Alia R Warner, Vincent M Dehili, Angela I Canto, and Eric L Garland. 2015. Washing dishes to wash the dishes: brief instruction in an informal mindfulness practice. *Mindfulness* 6, 5 (2015), 1095–1103.

[21] William M Hartmann and Stephen T Neely. 1998. Signals, sound, and sensation.

[22] Hermann LF Helmholtz. 2009. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Cambridge University Press.

[23] James H Howard, Alice J O'toole, Raja Parasuraman, and Kevin B Bennett. 1984. Pattern-directed attention in uncertain-frequency detection. *Perception & Psychophysics* 35, 3 (1984), 256–264.

[24] James H Howard Jr, Alice J O'Toole, and Susan E Rice. 1986. The role of frequency versus informational cues in uncertain frequency detection. *The Journal of the Acoustical Society of America* 79, 3 (1986), 788–791.

[25] Changshuo Hu, Qiang Yang, Yang Liu, Tobias Röddiger, Kayla-Jade Butkow, Mathias Ciliberto, Adam Luke Pullin, Jake Stuchbury-Wass, Mahbub Hassan, Cecilia Mascolo, et al. 2025. A survey of earable technology: Trends, tools, and the road ahead. *arXiv preprint arXiv:2506.05720* (2025).

[26] Ronald Hübner and Ervin R Hafter. 1995. Cuing mechanisms in autitory signal detection. *Perception & psychophysics* 57, 2 (1995), 197–202.

[27] Anna Lena Knoll, Tristan Barrière, Rosalie Weigand, Thomas Jacobsen, Helmut Leder, and Eva Specker. 2024. Experiencing beauty in everyday life. *Scientific reports* 14, 1 (2024), 9463.

[28] Sen M Kuo, Sohini Mitra, and Woon-Seng Gan. 2006. Active noise control system for headphone applications. *IEEE Transactions on Control Systems Technology* 14, 2 (2006), 331–335.

[29] Mats B Küssner and Daniel Leech-Wilkinson. 2014. Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing paradigm. *Psychology of Music* 42, 3 (2014), 448–469.

[30] Nelson Posse Lago and Fabio Kon. 2004. The quest for low latency. In *ICMC*.

[31] Hsiao Mun Lee, Ziyu Bai, Yenn Shen Ho, Jun Xiong Soh, and Heow Pueh Lee. 2020. Effect of music from headphone on pedestrians. *Applied acoustics* 169 (2020), 107485.

[32] Jingjin Li, Karen Anne Cochrane, and Gilly Leshed. 2024. Beyond Meditation: Understanding Everyday Mindfulness Practices and Technology Use Among Experienced Practitioners. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–29.

[33] Jingjin Li, Nayeon Kwon, Huong Pham, Ryun Shim, and Gilly Leshed. 2023. Co-designing magic machines for everyday mindfulness with practitioners. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1630–1647.

[34] Richard Lichenstein, Daniel Clarence Smith, Jordan Lynne Ambrose, and Laurel Anne Moody. 2012. Headphone use and pedestrian injury and death in the United States: 2004–2011. *Injury prevention* 18, 5 (2012), 287–290.

[35] Louise Lu, Jennifer Aaker, Yu Ding, and Szu-chi Huang. 2026. Tracking Beauty: How Attending to Beautiful Experiences Increases Intentions to Preserve Memories. (2026). Preprint available on ResearchGate. doi:10.13140/RG.2.2.27425.77921/1

[36] Teemu Mäki-Patola and Perttu Hämäläinen. 2004. Latency tolerance for gesture controlled continuous sound instrument without tactile feedback. In *ICMC*. Citeseer.

[37] Adam Moore and Peter Malinowski. 2009. Meditation, mindfulness and cognitive flexibility. *Consciousness and cognition* 18, 1 (2009), 176–186.

[38] Brian CJ Moore. 2012. *An introduction to the psychology of hearing*. Brill.

[39] Brian CJ Moore, Brian R Glasberg, and Morten L Jepsen. 2009. Effects of pulsing of the target tone on the audibility of partials in inharmonic complex tones. *The Journal of the Acoustical Society of America* 125, 5 (2009), 3194–3204.

[40] Brian CJ Moore, Brian R Glasberg, and Andrew J Oxenham. 2012. Effects of pulsing of a target tone on the ability to hear it out in different types of complex sounds. *The Journal of the Acoustical Society of America* 131, 4 (2012), 2927–2937.

[41] Mark B Neider, Jason S McCarley, James A Crowell, Henry Kaczmarski, and Arthur F Kramer. 2010. Pedestrians, vehicles, and cell phones. *Accident Analysis & Prevention* 42, 2 (2010), 589–594.

[42] Mia Huong Nguyen, Moritz Alexander Messerschmidt, Chitralekha Gupta, and Suranga Nanayakkara. 2025. GutIO: Toward Sensing and Inducing Gut Feelings with Abdominal Sounds. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.

[43] Lars Nieradzik. 2025. SwiftF0: Fast and Accurate Monophonic Pitch Detection. *arXiv preprint arXiv:2508.18440* (2025).

[44] Ted Painter and Andreas Spanias. 2002. Perceptual coding of digital audio. *Proc. IEEE* 88, 4 (2002), 451–515.

[45] Gyuseok Park, Woohyeong Cho, Kyu-Sung Kim, and Sangmin Lee. 2020. Speech enhancement for hearing aids with deep learning on environmental noises. *Applied Sciences* 10, 17 (2020), 6077.

[46] Diane K Reibel, Jeffrey M Greeson, George C Brainard, and Steven Rosenzweig. 2001. Mindfulness-based stress reduction and health-related quality of life in a heterogeneous patient population. *General hospital psychiatry* 23, 4 (2001), 183–192.

[47] Virginia M Richards and Donna L Neff. 2004. Cuing effects for informational masking. *The Journal of the Acoustical Society of America* 115, 1 (2004), 289–300.

[48] Henning Schepker, Florian Denk, Birger Kollmeier, and Simon Doclo. 2020. Acoustic transparency in hearables—Perceptual sound quality evaluations. *Journal of the Audio Engineering Society* 68, 7/8 (2020), 495–507.

[49] Robert S Schlauch and Ervin R Hafter. 1991. Listening bandwidths and frequency uncertainty in pure-tone signal detection. *The Journal of the Acoustical Society of America* 90, 3 (1991), 1332–1339.

[50] Andreas Schmid, Maria Ambros, Johanna Bogon, and Raphael Wimmer. 2024. Measuring the Just Noticeable Difference for Audio Latency. In *Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures*. 325–331.

[51] Peter Schneider and Martina Wengenroth. 2009. The neural basis of individual holistic and spectral sound perception. *Contemporary music review* 28, 3 (2009), 315–328.

[52] Catherine Semal and Laurent Demany. 1993. Further evidence for an autonomous processing of pitch in auditory short-term memory. *The Journal of the Acoustical Society of America* 94, 3 (1993), 1315–1322.

[53] Rebecca Shankland, Damien Tessier, Lionel Strub, Aurélie Gauchet, and Céline Baeyens. 2021. Improving mental health and well-being through informal mindfulness practices: An intervention study. *Applied Psychology: Health and Well-Being* 13, 1 (2021), 63–83.

[54] Sheng Shen, Nirupam Roy, Junfeng Guan, Haitham Hassanieh, and Romit Roy Choudhury. 2018. MUTE: Bringing IoT to noise cancellation. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. 282–296.

[55] Shan Shu and Hui Ma. 2020. Restorative effects of urban park soundscapes on children's psychophysiological stress. *Applied Acoustics* 164 (2020), 107293.

[56] Michael Speca, Linda E Carlson, Eileen Goodey, and Maureen Angen. 2000. A randomized, wait-list controlled clinical trial: the effect of a mindfulness meditation-based stress reduction program on mood and symptoms of stress in cancer outpatients. *Psychosomatic medicine* 62, 5 (2000), 613–622.

[57] John A Swets and Susan T Sewall. 1961. Stimulus vs response uncertainty in recognition. *The Journal of the Acoustical Society of America* 33, 11 (1961), 1586–1592.

[58] Magdalena Szumilas. 2010. Explaining odds ratios. *Journal of the Canadian academy of child and adolescent psychiatry* 19, 3 (2010), 227.

[59] Yujie Tao, Tan Gemicioglu, Sam Chin, Bingjian Huang, Jas Brooks, Sean Follmer, Pedro Lopes, and Suranga Nanayakkara. 2025. Toward Everyday Perceptual and Physiological Augmentation. In *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–4.

[60] Yujie Tao, Jingjin Li, Libby Ye, Andrew Zhang, Jeremy N Bailenson, and Sean Follmer. 2025. Audio Augmentation of Manual Interactions to Support Mindfulness. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 4 (2025), 1–27.

[61] Steve Temme, Pascal Brunet, and DB Keele Jr. 2009. Practical measurement of loudspeaker distortion using a simplified auditory perceptual model. In *Audio Engineering Society Convention 127*. Audio Engineering Society.

[62] Ernst Terhardt. 1979. Calculating virtual pitch. *Hearing research* 1, 2 (1979), 155–182.

[63] Mari Tervaniemi, Lauri Janhunen, Stefanie Kruck, Vesa Putkinen, and Minna Huotilainen. 2016. Auditory profiles of classical, jazz, and rock musicians: Genre-specific sensitivity to musical sound features. *Frontiers in psychology* 6 (2016), 1900.

[64] Bandhav Veluri, Malek Itani, Justin Chan, Takuya Yoshioka, and Shyamnath Gollakota. 2023. Semantic hearing: Programming acoustic scenes with binaural hearables. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[65] 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.

[66] Hiroki Watanabe and Tsutomu Terada. 2023. Transparency Mode of Hearable Reduces Your Spatial Hearing: Evaluation and Cancelling Method to Restore Spatial Hearing. *IEEE Access* 11 (2023), 97952–97960.

[67] Arthur Welwood. 1980. Improvisation with found sounds. *Music Educators Journal* 66, 5 (1980), 72–77.

[68] David Wessel and Matthew Wright. 2002. Problems and prospects for intimate musical control of computers. *Computer music journal* 26, 3 (2002), 11–22.

[69] Bernard Widrow, John R Glover, John M McCool, John Kaunitz, Charles S Williams, Robert H Hearn, James R Zeidler, JR Eugene Dong, and Robert C Goodlin. 1975. Adaptive noise cancelling: Principles and applications. *Proc. IEEE* 63, 12 (1975), 1692–1716.

[70] Dr Cristina Rolim Wolffenbüttel and Henrique Pellin. 2020. Sound research and composition with everyday sounds. *Int. J. Dev. Res* 10 (2020), 34464–34471.

[71] Yael Zaltz, Eitan Globerson, and Noam Amir. 2017. Auditory perceptual abilities are associated with specific auditory experience. *Frontiers in Psychology* 8 (2017), 2080.

[72] Fan-Gang Zeng, Stephen Rebscher, William Harrison, Xiaoan Sun, and Haihong Feng. 2008. Cochlear implants: system design, integration, and evaluation. *IEEE reviews in biomedical engineering* 1 (2008), 115–142.

[73] Ke Zhang. 2018. Identifying effective informal mindfulness practices in daily activities. *Journal of International Buddhist Studies* 9, 2 (2018), 1–18.

[74] Kai Zhen, Mi Suk Lee, Jongmo Sung, Seungkwon Beack, and Minje Kim. 2020. Psychoacoustic calibration of loss functions for efficient end-to-end neural audio coding. *IEEE Signal Processing Letters* 27 (2020), 2159–2163.

[75] Eberhard Zwicker. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America* 33, 2 (1961), 248–248.

[76] Eberhard Zwicker and Hugo Fastl. 2013. *Psychoacoustics: Facts and models*. Vol. 22. Springer Science & Business Media.