Earmony: Orchestrating Adaptive Auditory Cardio-Respiratory Biofeedback on Commodity Earables

Lily Sheng*
Key Laboratory of Pervasive Computing, Ministry of Education, Department of Computer Science and Technology
Tsinghua University
Beijing, China
cqr24@mails.tsinghua.edu.cn

Ruiqi Hong*
School of Information Management
Wuhan University
Wuhan, China
rachel@whu.edu.cn

Xiyuxing Zhang*
Department of Computer Science and Technology
Tsinghua University
Beijing, China
zxyx22@mails.tsinghua.edu.cn

Yuntao Wang†
Department of Computer Science and Technology
Tsinghua University
Beijing, China
School of Computer Science and Technology
Qinghai University
Xining, Qinghai, China
yuntaowang@tsinghua.edu.cn

Ruotong Yu
Global Innovation Exchange (GIX) Institute
Tsinghua University
Beijing, China
ruotong@uw.edu

Xingjian Tian
College of Intelligence and Computing, Tianjin University
Tianjin, China
tianxj_0520@tju.edu.cn

Yuanchun Shi
Department of Computer Science and Technology
Tsinghua University
Beijing, China
Qinghai University
Xining, Qinghai, China
shiyc@tsinghua.edu.cn

Abstract

Commodity earbuds are becoming an increasingly promising platform for everyday health and wellness support. Building on their wearability and ubiquity, we present Earmony, an adaptive auditory cardio-respiratory biofeedback system that runs on off-the-shelf earbuds. Using only the built-in in-ear microphone, Earmony unobtrusively senses cardio-respiratory signals without requiring additional sensors. To support simultaneous sensing and audio playback, we develop a robust detection framework that preserves signal quality under concurrent audio output. Building on this capability, we implement a closed-loop system in which auditory feedback is dynamically modulated by the user's real-time physiological state rather than following a fixed pace. Specifically, musical chords guide breathing and adapt continuously to both breathing dynamics and heart rate variability (HRV), enabling responsive and personalized regulation. A user study ($N = 20$) shows that Earmony maintains significantly elevated HRV for longer durations than static systems, demonstrating the potential of seamless, earbud-based adaptive biofeedback.

CCS Concepts

• **Human-centered computing → Auditory feedback**.

Keywords

Cardio-Respiratory Biofeedback, Auditory Biofeedback, Physiological Sensing, Earables

*Equal Contribution.
†Corresponding author.

UIST '26, Detroit, MI, USA

ACM ISBN 979-8-4007-2856-3/2026/11
https://doi.org/10.1145/3830398.3830578

ACM Reference Format:
Lily Sheng, Ruiqi Hong, Xiyuxing Zhang, Yuntao Wang, Ruotong Yu, Xingjian Tian, and Yuanchun Shi. 2026. *Earmony*: Orchestrating Adaptive Auditory Cardio-Respiratory Biofeedback on Commodity Earables. 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/10.1145/3830398.3830578

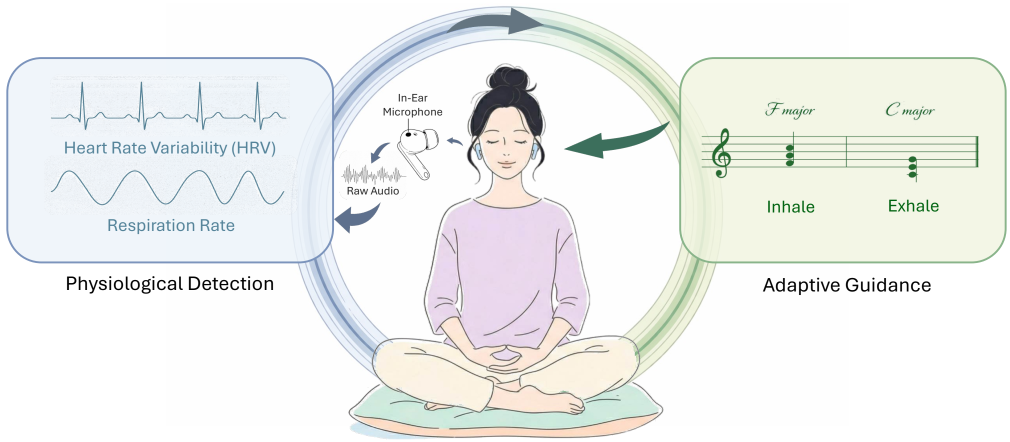

Figure 1: The closed-loop interaction between physiological detection and adaptive guidance of Earmony. Audio from the earbuds' in-ear microphone enables real-time detection of HRV and respiration, modulating musical feedback (F major/C major chords) to guide the user's breathing.

1 Introduction

Commodity earbuds have become a staple of modern life, offering a unique and promising platform for continuous health monitoring [8, 19, 43]. Since earbuds sit close to major vascular pathways and the upper respiratory tract, earables provide a high signal-to-noise ratio for detecting physiological data [8, 43, 56]. This proximity makes them ideal for seamless, always-on sensing and feedback systems.

Beyond raw signal capture, earables are increasingly applied to physiological monitoring for general wellness [35]. Central to this is Cardio-Respiratory Coupling (CRC), the natural synchronization between heart and breath rhythms [10, 15]. While high heart rate variability (HRV) serves as a primary indicator of a robust autonomic nervous system [16], stress and chronic conditions frequently disrupt this coupling [23]. To restore this balance, a true closed-loop system requires integrated cardio-respiratory biofeedback, which demands the simultaneous real-time sensing of both cardiac and respiratory signals. This ensures breathing guidance is not a fixed metronome, but a responsive intervention dynamically adapting to the user's real-time HRV and respiratory rate.

A significant technical barrier to integrated biofeedback is high-accuracy detection of physiological signals during active audio playback. This challenge stems from spectral masking, where the audio obscures subtle physiological vibrations captured by in-ear microphones. While cardiac signals are concentrated below 50 Hz and can be easily isolated via low-pass filtering, respiratory signals (150 Hz to 2000 Hz [25, 36]) overlap significantly with music and voice [64]. Consequently, while cardiac metrics can be reliably extracted through low-pass filtering, robust respiratory sensing requires a dedicated lane. Earmony addresses this by employing frequency carving to isolate a narrow 100 Hz band (150 Hz - 250 Hz) within the audio stream. This creates a specialized sensing lane that allows for the precise detection of breathing without perceptibly degrading the user's listening or relaxation experience.

Furthermore, most prior auditory systems treat breathing and HRV as separate channels, failing to provide a unified, closed-loop experience that adapts to both metrics. Earmony implements an adaptive biofeedback mechanism that encodes both HRV state and respiratory guidance simultaneously through the use of musical chords. Unlike discrete notes [21, 34, 62], these chords fuse harmonically related tones into a single auditory object, offering a rich information stream without added cognitive load. The system guides respiration through a dynamic amplitude envelope, while increasing harmonic complexity as a reward for improved HRV. As relaxation deepens, the audio continuously transitions from a single root tone to a rich, three-note major chord. This dual-factor approach ensures guidance continuously adapts to the user's physiological progress, providing a coherent acoustic reward for an improved autonomic state.

The effectiveness of Earmony's integrated approach was validated through a controlled study ($N = 20$). We systematically compared Earmony against only HRV biofeedback, only adaptive breathing guidance, or a static baseline. Results demonstrated that participants using Earmony maintained elevated HRV for significantly longer than other conditions. This confirms that the synergy of real-time HRV biofeedback and adaptive breathing guidance is necessary to prevent recovery decay and support a prolonged state of physiological balance.

Our primary contributions include:

(1) **Robust Sensing During Active Audio Playback**: A novel framework that isolates a narrow 100 Hz band (150–250 Hz) to detect breathing signals without degrading the user's perceived relaxation.
(2) **Integrated Cardio-Respiratory Auditory Cues**: A unified biofeedback signal that encodes both HRV state and respiratory guidance into musical chords.
(3) **Closed-Loop Auditory Adaptive Biofeedback**: Earmony implements a closed-loop auditory cardio-respiratory design providing simultaneous HRV and respiratory biofeedback alongside adaptive guidance that synchronizes with the user's live physiological state.

Table 1: Comparison of biofeedback systems across feedback type, commodity device, physiological modalities, and guidance adaptivity.

Systems	Feedback Type	Commodity		Modality		Guidance Adaptivity	Single Device
		Sensing	Feedback	RR	HRV		
Breathm [51]	Haptic	✗	✗	✓	✗	✓	✓
ViBreathe [57]	Haptic	✗	✗	✓	✗	✓	✓
Blum et al. [4]	Visual	✗	✓	✗	✓	✗	✗
Michela et al. [37]	Visual	✗	✗	✓	✓	✗	✗
BioFidget [29]	Visual	✗	✗	✓	✓	✓	✓
Breeze [48]	Visual	✓	✓	✓	✗	✗	✓
Breath of Life [7]	Visual	✓	✓	✓	✗	✗	✗
BreatheBuddy [41]	Visual	✓	✓	✓	✗	✓	✓
Yu et al. [60]	Visual	✓	✓	✓	✓	✓	✗
Yu et al. [58]	Auditory	✗	✓	✗	✓	✗	✗
Idrobo-Avila et al. [21]	Auditory	✗	✓	✗	✓	✗	✗
Unwind [59]	Auditory	✗	✓	✗	✓	✗	✗
Ours	Auditory	✓	✓	✓	✓	✓	✓

2 Related Work

2.1 Cardio-Respiratory Biofeedback

Cardiorespiratory coupling (CRC) represents the vital synchronization between cardiac and respiratory rhythms, primarily manifested through Respiratory Sinus Arrhythmia (RSA) [10, 15]. This natural fluctuation, where the heart accelerates during inhalation and decelerates during exhalation, serves as a critical indicator of autonomic health and is directly related to emotional regulation, cognitive performance, and vagal tone [15]. In practice, the robustness of this coupling is often measured via the root mean square of successive differences (RMSSD), a metric sensitive to the rapid, breath-induced changes in heart rate. While high RMSSD indicates a healthy, responsive system, CRC is frequently disrupted by physiological stress and chronic conditions such as hypertension. These states attenuate parasympathetic modulation and increase cardiovascular strain, ultimately leading to a significant reduction in heart rate variability [22, 40, 52, 55]. To restore CRC, biofeedback provides a mechanism that mirrors internal physiological processes back to the user, enabling conscious regulation of typically involuntary functions [14, 47]. HRV biofeedback (HRVB) externalizes heart rate variability, empowering users to regulate their automatic state through breath [27, 54].

To deliver biofeedback, prior systems have predominantly relied on visual and haptic modalities. Visual systems project physiological states onto screens or tangible interfaces [29, 48, 60], which effectively guide breathing, but require focal attention, limiting their use during daily tasks. To reduce visual load, haptic interfaces like ViBreathe [57] and Breathm [51] provide somatic guidance. However, they typically require custom-built, non-commodity hardware worn on specific body parts, creating friction for daily usage.

In contrast, auditory feedback presents a compelling and eye-free alternative [21, 58], which is also widely adopted in various cognitive regulation practices, including mindfulness and meditation [11, 20]. Since rhythmic auditory cues can provide a consistent temporal anchor that naturally entrains biological oscillators, auditory feedback enable users to stabilize their breathing and heart rate through intuitive synchronization [3]. Table 1 highlights that, although prior work has demonstrated the efficacy of auditory HRV biofeedback [21, 58, 59], existing systems largely rely on custom hardware or separate sensing and feedback components, rather than a single commodity device.

Yu et al. [60] successfully implemented adaptive cardio-respiratory biofeedback using commodity devices, yet their pipeline separated sensing (smartwatches and smart rings) from visual feedback (smartphones). This separation leaves open the question of how to achieve closed-loop sensing and biofeedback within a single device, rather than distributing those roles across multiple wearables. Leveraging earables for integrated auditory biofeedback highlights a unique opportunity to explore all-in-one paradigms that could streamline health interventions and broaden the reach of digital care.

2.2 Commodity Earbuds as Health Entry

Earables are rapidly evolving from audio peripherals into ubiquitous personal health gateways [43], a paradigm shift underscored by recent integration of hearing tests and heart rate tracking in Airpods [1]. Specifically for HRV biofeedback (HRVB), earbuds offer an ideal all-in-one platform by unifying continuous sensing and built-in audio playback.

From a sensing perspective, earables have emerged as a promising cardio-respiratory platform due to their unique anatomical positioning [8]. Positioned close to major vascular pathways and the upper respiratory tract, earables provide a high signal-to-noise ratio (SNR) for the acquisition of both cardiac and respiratory rhythms [8, 43, 56]. While dedicated sensors like PPG [13, 45, 50] and IMU [1, 42, 44] can capture cardio-respiratory metrics, they are not widely available on commodity earbuds. In contrast, the universal adoption of Active Noise Cancellation (ANC) has made the in-ear microphone a standard, always-on sensor. Researchers have demonstrated that in-ear microphones can be used to effectively capture cardiac and respiratory activities via acoustic signals [12, 30–32], scaling cardio-respiratory sensing to a much broader range of everyday commodity earbuds without requiring supplementary sensors.

[1] Apple Newsroom: Apple introduces AirPods 4 and a hearing health experience. https://www.apple.com/newsroom/2024/09/apple-introduces-airpods-4-and-a-hearing-health-experience-with-airpods-pro-2/

Beyond physiological sensing, earables emerge as an optimal platform for delivering auditory biofeedback. By providing high-fidelity audio directly into the ear canal, earbuds create an immersive and private acoustic environment that shields users from external disturbances, which is a critical requirement for effective cognitive and emotional regulation [61]. Consequently, users already widely adopt earbuds during practices such as yoga, meditation, and mindfulness to consume relaxing soundscapes and breathing instructions [24, 28].

Despite these advantages of sensing and auditory feedback, realizing an all-in-one system on earables for closed-loop HRVB introduces a severe technical challenge. The continuous audio playback required to guide the user actively masks the incredibly subtle physiological sounds of respiration and heartbeats within the ear canal. Consequently, unlocking integrated biofeedback on commodity earbuds demands a robust acoustic pipeline capable of isolating micro-physiological signals from concurrent audio playback.

2.3 Auditory HRV Biofeedback Systems

Breathing and HRV biofeedback have largely been explored as separate modalities, each with distinct auditory cue designs. Prior work on breathing biofeedback has explored diverse auditory designs to signal respiratory phases, including breath-like pink noise [53], synthetic breathing stimuli [34], naturalistic sounds such as wind-blown leaves [53], and musical mappings using harmonic chords [62] or melodic contours [34] to indicate inhalation and exhalation. Several systems layer multiple concurrent cues: pairing musical tones with synthetic breath sounds [34], or combining musical notes with pink-noise breath sounds [53]. However, increasing the number of concurrent auditory cues may elevate cognitive load [53] and reduce auditory clarity.

Compared to breathing biofeedback, auditory HRV biofeedback remains relatively underexplored. Yu et al. [58] proposed real-time auditory displays that mapped HRV to temporal sound variations and demonstrated training effects comparable to visual biofeedback. Idrobo-Avila et al. [21] adopted 24 common harmonic musical intervals (pairs of notes played simultaneously) to modulate HRV. *Unwind* [59] integrated HRV-responsive nature sounds with a relaxation-oriented soundscape, incorporating wind-like sounds as implicit breathing guidance. However, none of these systems employed real-time respiratory sensing, making the breathing support non-adaptive and precluding true closed-loop guidance.

These gaps point to a key design challenge: modulating breathing pace adaptively while simultaneously communicating both respiratory and cardiac cues, as excessive concurrent auditory signals may increase cognitive load and undermine intervention effectiveness.

3 Sensing Design

3.1 Frequency Carving

To facilitate the real-time detection of respiratory signals during active audio playback, frequency carving is employed to mitigate spectral masking. This phenomenon occurs when a high-amplitude signal, such as music or voice playback, obscures lower-amplitude physiological signals within the same frequency band. While cardiac-related acoustic components are concentrated primarily below 50 Hz and remain largely unaffected by standard audio playback, respiratory acoustic signals occupy a much broader range [64]. The acoustic signal of breathing captured through an in-ear microphone typically lies in the range of 150 Hz to 2000 Hz [25, 36], which overlaps substantially with the frequency spectrum of most music [64]. Consequently, robust respiratory sensing requires the use of frequency carving. We select the lower end and remove frequencies from 150 Hz to 250 Hz from the audio playback, creating a narrow 100 Hz band from which we can detect breathing sounds.

To quantitatively evaluate the effectiveness of frequency carving, we conducted a pilot study with 4 participants (2 females, 2 males, aged 21-23). While maintaining normal breathing, participants listened to three representative meditative audio tracks under two conditions: Normal Audio and Audio with Frequencies 150-250 Hz Removed. Condition order was counterbalanced across trials. With frequency carving, the RR estimation pipeline (detailed in Section 3.4) achieved MAE = 1.42 br/min and MAPE = 22.11%. Without frequency carving, the performance degraded to MAE = 2.52 br/min and MAPE = 30.33%. For the overall study, the heart rate metrics achieved HR MAE = 8.09 bpm, MAPE = 9.66% with IBI MAE = 84.52 ms, MAPE = 11.80%. These quantitative results demonstrate that frequency carving substantially reduces respiration estimation error, enabling reliable respiration sensing during active audio playback.

3.2 User Experience

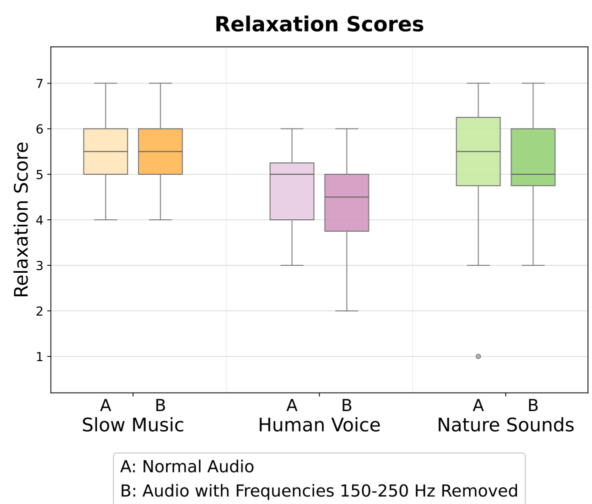

Figure 2: Comparison of subjective relaxation scores between unprocessed audio and processed audio with 150–250 Hz frequency carving. Box plots represent the median and IQR.

3.2.1 Study Procedure. To evaluate the subjective impact of frequency carving, we conducted a within-subject study with 12 participants (8 females, 4 males, aged 21 - 30). Participants listened to three representative categories of meditative and breathing exercise audio: Slow Music, Human Voice, and Nature Sounds. Within each category, we played two acoustic variants: (A) Normal Audio and (B) Audio with Frequencies 150-250 Hz Removed. The order of the three audio conditions was counterbalanced, with the order of A and B conditions randomized. Following each audio, participants rated their perceived state of relaxation on a 7-point Likert scale, ranging from 1 (Not at all Relaxed) to 7 (Extremely Relaxed). This study was approved by our Institutional Review Board. Informed written consent was obtained from participants in the study before any study procedures began.

3.2.2 Results. To determine if the spectral modification negatively impacted the perceived relaxation of the participants, a one-sided Wilcoxon signed-rank test was performed for each audio category. The analysis compared the unprocessed audio (A) against the processed audio (B) under the hypothesis that A>B (i.e., that filtering would reduce relaxation).

Results did not support the hypothesis that filtering would reduce relaxation ($p > 0.05$ for all three categories), indicating no statistically significant difference in relaxation scores between unprocessed (A) and processed (B) audio (Figure 2). These findings suggest that the removal of the 150–250 Hz frequency band does not significantly diminish the subjective relaxation quality of mindfulness audio, supporting the viability of frequency carving.

3.3 Data Collection

To develop and validate Earmony's cardiac and respiratory sensing algorithms, we conducted a controlled data collection study.

Participants. We recruited 18 participants (15 females, 3 males, aged 21–30). This study was approved by our Institutional Review Board, and informed written consent was obtained from all participants prior to any study procedures.

Device Setup. Ground-truth cardiac metrics and respiratory rate were recorded using a VIVALINK ECG patch and an HKH-11C piezoelectric respiratory belt, respectively. Each participant simultaneously wore two earables: a commodity Honor Earbud equipped with an in-ear microphone, and a reference PPG/IMU earable. To mitigate potential ear-side effects, half of the participants wore the in-ear microphone earable on the left ear and the reference device on the right, with the remaining half wearing them on opposite sides. All devices were synchronized using a common percussive tap signal captured across all recording channels.

Procedure. Each session began with a 6-minute resting period under normal breathing. Participants then completed four counterbalanced sessions, each preceded by a standardized Trier Social Stress Test (TSST) arithmetic task [2] to induce a consistent stress state prior to intervention. The four sessions comprised: (1) a 6-minute *deep breathing* session with no audio content, and (2–4) three 3-minute *audio biofeedback* sessions featuring slow music, nature sounds, and vocal guidance, respectively. The 6-minute duration of both the resting and deep breathing sessions ensured sufficient data for model training, as both were used as training data for the respiratory sensing algorithm. Except for the vocal guidance condition, which provided explicit verbal breathing cues, all other sessions used a visual pacer (an animated expanding-contracting circle labeled *Inhale/Exhale*) [38] to ensure participants maintained the target breathing pace. Each participant was randomly assigned a fixed breathing rate of 6, 7, 8, or 9 BPM to ensure broad respirator coverage. Session ordering across the three audio conditions was counterbalanced using a balanced Latin square, with deep breathing sessions alternated to yield six total orderings. Participants were not instructed to restrict their movements during the intervention, allowing naturally occurring motions such as head tilts, posture shifts, and breathing adjustments to occur.

3.4 Sensing Algorithm

3.4.1 HRV Detection. To detect cardiac metrics, we first process the raw audio to a filtered signal and then use peak detection.

Preprocessing. The raw audio signal acquired from the in-ear microphone is processed to isolate cardiac-related mechanical vibrations (Figure 3). Given that the primary cardiac components captured within the ear canal reside in the low-frequency spectrum below 50 Hz [6, 64], we first apply a third-order Butterworth bandpass filter with cutoff frequencies of 8 Hz and 50 Hz. This effectively suppresses high-frequency ambient noise and low-frequency motion drift. To better resolve the temporal structure of each heartbeat, we compute the analytic signal via the Hilbert transform [63]. The resulting magnitude provides an amplitude envelope that transforms oscillatory vibrations into a unipolar representation, making individual beats more distinguishable. This envelope is further smoothed using a low-pass Butterworth filter (3 Hz cutoff) to eliminate minor fluctuations while preserving the underlying heart rate dynamics necessary for stable peak detection.

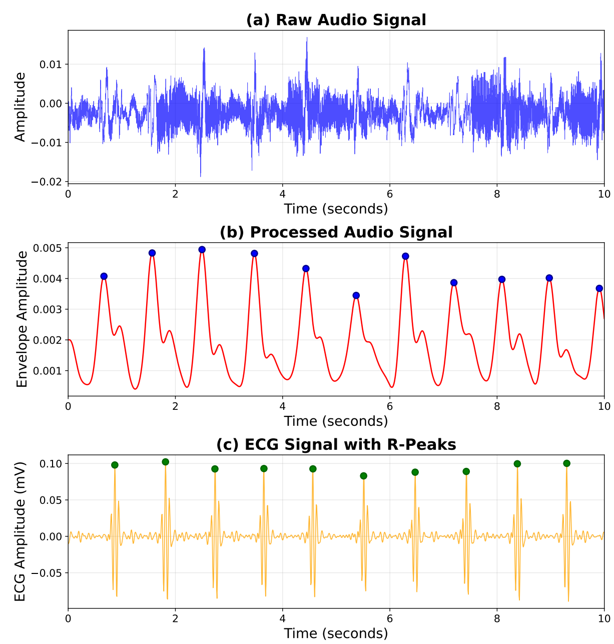

Figure 3: Comparative visualization of (a) raw audio from the in-ear microphone, (b) the processed audio signal, and (c) ground truth ECG reference.

Peak Detection. Heartbeat identification is performed using a template-based matched filtering approach, which enhances robustness against signal variability. We first identify a set of initial candidate peaks from a short segment of the smoothed envelope by applying amplitude prominence and physiological constraints (40 to 180 bpm). A representative heartbeat waveform is then extracted around the most prominent candidate to serve as a template of the user's unique cardiac-induced vibration morphology. By correlating this template with the full preprocessed signal, we generate a similarity index that highlights segments matching the learned heartbeat shape. Final peak locations are extracted from this matched filter output, using a minimum temporal distance constraint to prevent physiologically implausible detections. These peak locations define the heartbeat events used to calculate inter-beat intervals (IBI) for subsequent HRV analysis.

HRV Metrics. To quantify heart rate variability, we compute two standard time-domain metrics: SDNN and RMSSD. SDNN (Standard Deviation of Normal-to-Normal intervals) is calculated as the standard deviation of the IBI series in milliseconds, reflecting the overall cyclic components of cardiac variability. RMSSD is derived by taking the square root of the mean of the squared differences between adjacent IBI values.

The system achieved an overall Mean Absolute Error (MAE) for SDNN of 37.58±41.53 ms. For RMSSD, the system demonstrated a higher relative precision with an MAE of 29.14±26.83 ms. We selected RMSSD as the primary HRV metric for our intervention logic. This metric is particularly sensitive to high-frequency heart rate fluctuations and provides a more robust estimate of parasympathetic activity. Based on the experimental results, it also provides a more stable representation of beat-to-beat changes in the presence of the acoustic micro-jitter inherent to in-ear heart sound sensing.

Cardiac Signal Quality. The signal quality of this system was evaluated by computing HR and inter-beat interval IBI estimation errors. Specifically, MAE and MAPE were calculated using ECG-derived measurements as ground truth. In addition, performance was compared against a reference IMU/PPG earbud under the same conditions using methods to calculate cardiac metrics from Sah et al. [46] using the PPG signals. Earmony consistently achieved lower estimation errors than the IMU/PPG earbud across all segments (Table 2). These results suggest that the in-ear audio signal provides a stable representation of cardiac activity to support beat-to-beat analysis.

Table 2: Cardiac metrics estimation accuracy of Earmony (in-ear microphone) versus a reference PPG/IMU earbud [46], across three meditation audio conditions: G1 (slow music), G2 (human voice), and G3 (nature sounds).

Metric	Seg.	Earmony		PPG/IMU Earbud	
		MAE	MAPE(%)	MAE	MAPE(%)
HR (bpm)	G1	5.38	6.98	16.09	18.80
	G2	2.75	3.98	16.89	20.58
	G3	6.79	8.66	13.50	15.34
IBI (ms)	G1	64.11	8.35	169.23	23.98
	G2	37.43	4.31	183.11	25.15
	G3	79.91	10.54	137.38	20.01

3.4.2 Respiratory Rate Detection. To estimate respiratory rate (RR) from in-ear audio, we developed a machine learning pipeline integrated with adaptive signal processing, using a respiratory belt as ground truth.

Feature Extraction and Model Training. Based on the frequency carving method in Section 3.1 and the user study in Section 3.2, breathing-related airflow acoustics are concentrated in the 150–250 Hz band without perceptibly disrupting user relaxation. We therefore bandpass-filtered the 4 kHz in-ear audio to this range. We extracted 13-dimensional log-mel spectrograms with 100 ms frames and 50 ms hop, then appended first- and second-order temporal deltas, yielding 39-dimensional feature vectors at 20 frames/s. We selected log-mel over MFCCs based on empirical comparison on our dataset. Among four candidate classifiers (SGD, SVC, XGBoost, Random Forest), Random Forest with balanced class weighting achieved the best generalization and was adopted for all subsequent experiments.

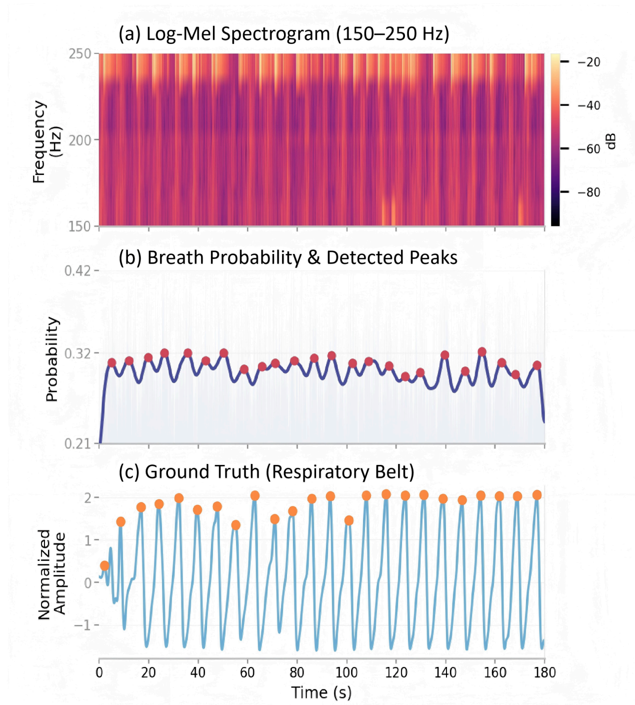

Figure 4: Respiratory rate estimation pipeline. (a) Log-Mel spectrogram of the filtered acoustic signal (150–250 Hz). (b) Breath probability predicted by the classifier (light purple area) and the smoothed breath envelope (deep indigo line), with red dots indicating the detected peaks. (c) Ground truth obtained from the respiratory belt.

From Probability to RR. BreathPro [18] performs end-to-end multi-class phase classification, using under-nose microphone signals as ground truth. With belt-based labels, precise inhale/exhale boundary annotation proved unreliable due to phase offsets introduced by abdominal displacement measurement. Instead, we treat the per-frame posterior probability as a continuous breath-likelihood signal. We apply two-stage smoothing: a 3 s moving-average window to suppress high-frequency noise, followed by Gaussian filtering (σ = 20 frames) to extract the slow respiratory envelope. Breath cycles are then located via adaptive peak detection with a locally normalized threshold, and RR is derived from the median inter-peak interval. The complete signal processing workflow and a representative detection result are visualized in Figure 4.

Table 3: Respiratory rate estimation accuracy of Earmony (in-ear microphone) versus a reference PPG/IMU earbud (RR via IMU [1]) across three meditation audio conditions: G1 (slow music), G2 (human voice), and G3 (nature sounds).

Metric	Seg.	Earmony		PPG/IMU Earbud	
		MAE	MAPE(%)	MAE	MAPE(%)
RR (br/min)	G1	0.65	9.00	1.52	15.86
	G2	0.95	12.41	1.43	16.27
	G3	0.87	11.14	2.27	19.51

Evaluation. To evaluate generalization to unseen users, we validated the pipeline under Leave-One-Subject-Out (LOSO) cross-validation ($N = 18$). The classifier was trained on resting and deep-breathing segments from 17 participants and tested on the audio-playback phase of the held-out participant, ensuring no leakage of subject-specific features or intervention-phase audio during training. Full per-segment results are reported in Table 3. The pipeline achieved a mean MAE of 0.82 ± 0.77 breaths/min and MAPE of 10.85 ± 9.32% across all three meditation audio conditions, consistently outperforming the reference PPG/IMU earbud (RR derived from the IMU channel [1]; mean MAE 1.75 ± 2.46 br/min; MAPE 17.27 ± 14.23%). During live intervention, the same pipeline runs on a rolling 30 s buffer updated every 5 s, with RR further stabilized via exponential moving average (EMA) across updates to suppress between-window variability.

4 Earmony System

4.1 Auditory Biofeedback Design

Earmony delivers biofeedback through real-time auditory cues encoding both HRV state and respiratory guidance. Prior biofeedback work has employed harmonic chords [62] or single notes [34, 58] as respiratory cues. Unlike discrete single-note sequences, chords fuse harmonically related tones into a single auditory perceptual object [33], offering greater richness without added cognitive load. We therefore adopt chords as our biofeedback carrier. We select major chords for their association with positive valence and, at slow tempo, calmness [26]. Voicings are drawn from the middle register (F4-A4-C5 for inhalation; C4-E4-G4 for exhalation), where tones in this register are perceived as warm and stable, facilitating relaxation [9].

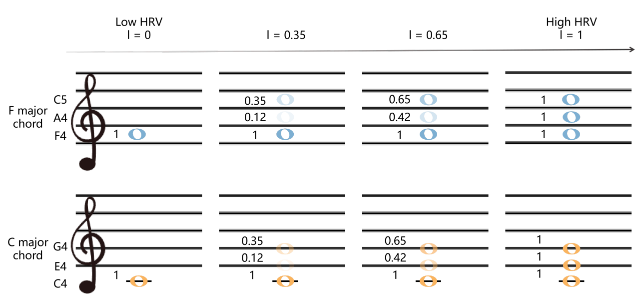

Figure 5: Dynamic mapping of HRV progress (l) to chord complexity. As (l) increases from 0 (Low HRV) to 1 (High HRV), F major and C major chords gain harmonic richness through the addition of notes with varying intensity weights.

HRV progress is encoded as chord complexity as l, where low HRV is $l = 0$ and high HRV is $l = 1$ (Figure 5). As the user's RMSSD approaches the session target, harmonic voices are progressively layered: starting from the root tone alone, the perfect fifth is added with amplitude scaled by l, and the major third by l^2, such that only the root tone is audible at low HRV levels, while a full three-note chord emerges as relaxation deepens. Since chord tones share integer-ratio frequency relationships, the added voices blend into a richer perceptual object rather than competing streams [33], functioning as an acoustic reward for improved autonomic state.

Respiratory guidance is embedded in the chord's dynamic envelope (Figure 6): an F-major chord modulates with a Gaussian amplitude profile during inhalation, followed by a C-major chord during exhalation [62], providing intuitive, non-verbal breath-pacing cues. Chords are rendered with a warm pad timbre to promote relaxation.

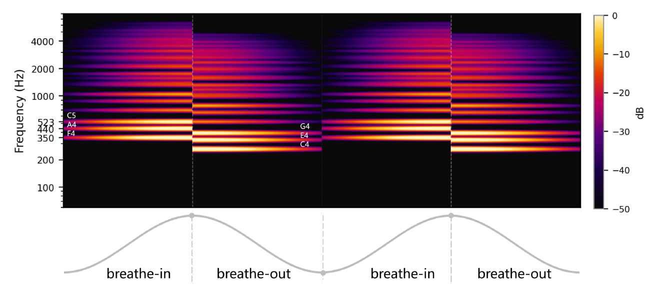

Figure 6: Simulated frequency spectrogram of the breath-pacing audio over two cycles. The primary bands (260–530 Hz) represent chord fundamentals for inhalation (F4–A4–C5) and exhalation (C4–E4–G4), while higher-frequency components correspond to the harmonic overtones.

4.2 Dual-Factor Adaptive Biofeedback

Building on real-time HRV (RMSSD) and respiration rate (RR) sensing (Section 3), Earmony implements a dual-factor adaptive biofeedback mechanism. After a 20 s calibration phase, the system operates a closed-loop controller with asynchronous sliding windows at multiple timescales. The controller prioritizes real-time respiratory synchronization before applying HRV-driven adaptation, ensuring stable and non-oscillatory regulation.

4.2.1 Initial Calibration and Parameter Initialization. Since Earmony does not pre-estimate each user's resonant frequency, the initial guidance rate must be safely initialized. Thus, Earmony performs a 20 s calibration to synchronize the system with the user's spontaneous breathing. By detecting the initial peak-to-peak intervals (T_{init}) of the smoothed respiratory probability curve, Earmony estimates the user's resting respiration rate during calibration:

$$RR_{init} = 60/T_{init} \tag{1}$$

To ensure a safe yet effective starting point, the initial guidance rate GR_{target} is initialized by clipping RR_{init} to a conservative range of 8.0–10.0 bpm.

4.2.2 *Hierarchical Adaptive Control Logic.*

Earmony employs an asynchronous multi-rate buffer: a 30 s window for respiration rate with a 5 s update to drive the fast-loop entrainment, and a 48 s window for HRV with a 2 s update to provide a stable, trend-based signal for slow-loop optimization.

Fast-loop: Respiration-driven Entrainment. Respiration serves as the fast-loop entrainment to enforce synchronization between the user and the system. For each window t, the controller calculates the entrainment error e_t between the guidance rate GR_t and the user's actual respiration rate RR_t:

$$e_t = |RR_t - GR_t| \tag{2}$$

A window is considered locally synchronized if $e_t < 2.0$ bpm, denoted by the indicator $\text{Sync}_t = \mathbb{I}(e_t < 2.0)$. The system defines a global synchronization state S_{synced} through a majority vote over the last three estimation cycles (45 s):

$$S_{synced} = \mathbb{I}\left(\sum_{i=t-2}^{t} \text{Sync}_i \geq 2\right) \tag{3}$$

where $\mathbb{I}(\cdot)$ is the indicator function. If a significant divergence occurs ($e_t > 2.0$ bpm), the controller executes a catch-up mechanism to adjust GR by $\Delta = 0.5$ bpm toward RR_t, ensuring the guidance remains reachable.

Slow-loop: HRV-driven Optimization and Gating. HRV acts as the slow-loop optimization to maximize autonomic engagement. To mitigate transient noise, we compute an Exponential Moving Average (EMA) of RMSSD to represent the long-term HRV trend:

$$HRV_{ema}(t) = \alpha \cdot RMSSD_t + (1 - \alpha) \cdot HRV_{ema}(t-1) \tag{4}$$

where $\alpha = 0.2$. A Sync-first Gating Mechanism ensures that HRV-driven modulation occurs only when $S_{synced} = 1$, preventing unstable updates during poor entrainment.

When gated, the system calculates the physiological deviation $\Delta HRV = RMSSD_t - HRV_{ema}(t)$. To promote gradual slowing of breath, we set asymmetric sensitivity thresholds for decreasing ($\theta_{\downarrow} = 0.1 \cdot HRV_{ema}$) and increasing ($\theta_{\uparrow} = 0.2 \cdot HRV_{ema}$) the rate:

$$GR_{t+1} = \begin{cases} GR_t - \Delta & \text{if } \Delta HRV > \theta_{\downarrow} \\ GR_t + \Delta & \text{if } \Delta HRV < -\theta_{\uparrow} \\ GR_t & \text{otherwise} \end{cases} \tag{5}$$

4.2.3 *Dynamic Progression and Termination.*

To sustain engagement and avoid habituation, Earmony dynamically scales physiological targets as the user progresses, and autonomously fades out once relaxation is achieved.

Adaptive Goal Scaling. To sustain user engagement, Earmony employs a ladder mechanism that dynamically scales physiological goals based on real-time performance.

The system modulates a goal multiplier $M_t \in [1.0, 1.2]$, defining the current target as $HRV_{target}(t) = HRV_{base} \cdot M_t$. To ensure the intervention is appropriately challenging, M_t increases only when $RMSSD_t \geq HRV_{target}$. The update step η is adaptively scaled by the user's "overshoot" to accelerate progression during rapid relaxation:

$$\text{overshoot} = \frac{RMSSD_t - HRV_{target}(t)}{HRV_{target}(t)} \tag{6}$$

$$\eta = \text{clip}(0.05 \cdot \text{overshoot},\ 0.005,\ 0.03) \tag{7}$$

This adaptive scaling accelerates progression under strong physiological responses while maintaining stability through bounded updates. The progression $M_{t+1} = M_t + \eta$ is then mapped to l, which controls the chord complexity described in Section 4.1.

Fade-out Phase. The system enters a fade-out phase when both conditions are sustained for at least 20 s: (1) $4.5 \leq RR \leq 6.5$ bpm [49], and (2) $M_t \geq 1.20$. Additionally, a 300 s timeout triggers the same procedure to bound session duration. Upon triggering, the audio volume V is gradually attenuated at each breathing cycle to gradually withdraw external guidance: $V_{t+1} = \max(0.3, V_t - 0.1)$. This facilitates a transition from guided regulation to self-sustained breathing.

4.3 System Implementation

Earmony runs as a Python application on a host laptop. The commodity Honor Earbud streams in-ear microphone audio via Bluetooth to the laptop, where the sensing pipeline (Section 3) continuously estimates RMSSD and RR in real time. These physiological estimates are passed to the dual-factor adaptive controller (Section 4.2), which computes the current guidance rate and HRV progress level l. The biofeedback audio is then synthesized in real time using a Python-based audio engine SCAMP, modulating chord complexity and breath-pacing envelopes as described in Section 4.1, and rendered through the same earable to close the feedback loop.

5 User Study: Intervention Evaluation

To evaluate Earmony's effectiveness, we conduct a controlled user study focusing on physiological recovery following a TSST task. We systematically compare our closed-loop cardio-respiratory biofeedback system against non-adaptive and single-modality baselines to quantify its impact on HRV recovery.

5.1 Participants

The study included 20 participants (12 females, 8 males), aged 19-33 years (M = 23.20, SD = 3.33). Participants were recruited through word-of-mouth and snowball sampling. This study was approved by our Institutional Review Board. Informed written consent was obtained from participants in the study before any study procedures began.

5.2 Study Design and Procedure

5.2.1 *Study Design.*

Our system integrates both HRV biofeedback and adaptive breathing guidance. To separate the two factors and examine how each contributes to the overall system, we conducted an ablation study using a 2×2 factorial design (Table 4). Participants were randomly assigned to one of the four experimental conditions.

Table 4: Comparison of experimental groups across HRV biofeedback and guidance adaptivity parameters.

Condition	HRV Biofeedback	Adaptive Guidance (RR)
Fixed-None	✗	✗
Fixed-HRV	✓	✗
Adaptive-RR	✗	✓
Adaptive-Both	✓	✓

- **Fixed-None (Baseline)**: The baseline condition with a fixed breathing guidance of 6 bpm and no HRV biofeedback. The chord complexity level is fixed ($l = 0.5$).
- **Fixed-HRV**: This condition has real-time HRV biofeedback, but has a fixed breathing guidance of 6 bpm.
- **Adaptive-RR**: This condition has adaptive breathing guidance, but no HRV biofeedback (chord fixed at ($l = 0.5$).
- **Adaptive-Both**: This is the fully-integrated system, Earmony, with adaptive breathing guidance and HRV biofeedback.

5.2.2 Study Procedure. Upon arrival, participants were fitted with the earbuds, followed by a 2-minute recording to establish a resting HRV baseline. Stress was then induced using the standardized Trier Social Stress Test (TSST) arithmetic task [2]. To ensure a consistent physiological starting point for the recovery phase across all participants, the stress induction period concluded only once the participant's HRV dropped to 85% of their initial baseline. Participants then engaged in a 6-minute guided breathing session corresponding to their randomly assigned experimental condition. Following the intervention, participants completed three standardized questionnaires: the User Engagement Scale - Short Form (UES-SF), specifically the Focused Attention and Reward subscales, to evaluate user experience and engagement quality [39]; the NASA Task Load Index (NASA-TLX) to assess perceived cognitive and physical workload [17]; and the System Usability Scale (SUS) to measure the technical usability of the intervention system [5]. Each questionnaire was scored on a 7-point Likert scale.

5.3 Results

5.3.1 Evaluation Metrics.

- **Average ΔHRV Time Series**: The temporal progression of the mean change in HRV across the 6-minute intervention. Each participant's data is synchronized at $T = 0$, representing the specific moment their HRV dropped to 85% of their resting baseline. By aligning participants to this consistent physiological state rather than a fixed clock time, this metric accurately visualizes the dynamic rate of recovery and the trend of physiological balance for each condition.
- **Max ΔHRV**: The maximum increase in HRV from the participant's baseline, representing the peak physiological relaxation response achieved during intervention.
- **Time Above 120% Baseline**: This calculates the cumulative duration, in seconds, during which the participant's HRV remained significantly elevated above their initial state. This provides a measure of sustained effectiveness of the intervention.
- **Average ΔHRV**: The global mean change in HRV for each condition during intervention. It provides a measure of each condition's overall impact on recovery from stress.

5.3.2 HRV Response Results. We analyze the HRV response results across all four conditions.

Average ΔHRV Time Series. The Adaptive-Both group exhibited a distinct and sustained upward trajectory throughout the session, separating from the other conditions within the first 60 seconds (Figure 7). While all groups showed an initial recovery spike in the first minute, participants found it significantly harder to maintain a high state of recovery using a static pacer or adaptive pacing alone. As the session progressed, the Fixed-None and Adaptive-RR groups trended back toward or below the $T = 0$ baseline after the initial 100 seconds. Similarly, the Fixed-HRV group showed moderate fluctuations but ultimately failed to sustain the high recovery levels achieved by the fully integrated system. These time-series results suggest that while various guidance modes can trigger an initial relaxation response, the synergy of real-time biofeedback and adaptive guidance is necessary to prevent recovery decay and support a prolonged state of physiological balance.

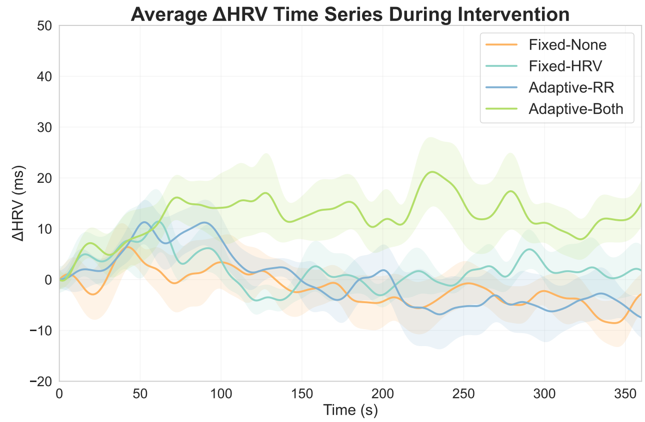

Figure 7: Average ΔHRV time series illustrating the physiological recovery trajectory over the 360-second post-stress window. The solid lines represent the mean, and the shaded areas represent the ± SEM for each group.

Max ΔHRV. The Adaptive-Both group achieved the highest HRV increase (M=35.4 ms, SD=22.3 ms), Figure 8. While pairwise Mann-Whitney U tests did not reveal statistically significant differences between the groups for this specific metric, the trend suggests that the integrated system effectively facilitates reaching high-magnitude relaxation states.

Time Above 120% Baseline. The sustainability of the relaxation effect was measured by the cumulative time participants maintained an HRV level significantly above their baseline (Figure 8). The Adaptive-Both group maintained this elevated state for significantly longer (M=248.0 s, SD=117.4 s) than the Fixed-None control group (M=68.0 s, SD=94.1 s; p=0.032) and the Adaptive-RR group (M=75.0 s, SD=78.7 s; p=0.032). No significant difference was found between the Fixed-None and Fixed-HRV (M=79.0 s, SD=90.5 s) conditions.

Average ΔHRV. The Adaptive-Both condition demonstrated the most substantial impact on physiological recovery (M=15.19 ms, SD=10.13 ms), Figure 8. This group achieved significantly higher recovery compared to the Fixed-None control group (M=-2.90 ms, SD=10.19 ms; p=0.023) and the Adaptive-RR group (M=-1.27 ms, SD=8.23 ms; p=0.022). The Fixed-HRV group showed a modest positive trend (M=2.45 ms, SD=7.78 ms). However, the difference compared to the control was not statistically significant.

5.3.3 Subjective Evaluation: Engagement, Workload, and Usability. Beyond physiological metrics, user experience was evaluated using standardized scales for engagement, cognitive load, and usability

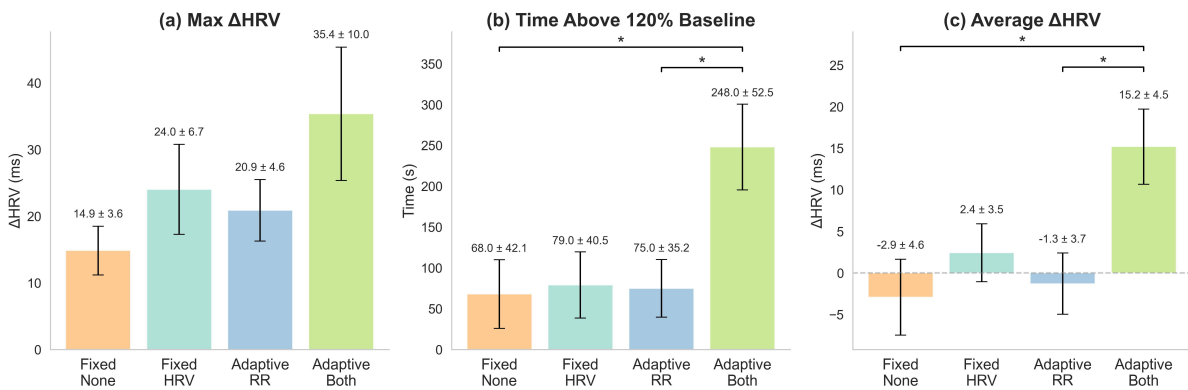

Figure 8: Comparison of HRV metrics across four intervention conditions. Bars plot the mean and standard error of the mean (SEM). Asterisks denote statistically significant differences between pairs of conditions based on pairwise Mann-Whitney U tests (a) and (b), independent-samples t-tests (c) (*p < 0.05, **p < 0.01, ***p < 0.001).

(Table 5). Median engagement (UES-SF) was highest in the Fixed-HRV group (6.0, IQR 5.4–6.6) and lowest for Adaptive-RR (5.4, IQR 4.2–5.4). Perceived workload (NASA-TLX) remained consistently low across conditions, ranging from 2.33 (IQR 2.33–3.3) (Fixed-HRV) to 3.0 (IQR 2.5–4.1) (Adaptive-Both), indicating that adaptive pacing imposed no significant cognitive burden. System usability (SUS) scores were stable, with Fixed-HRV and Adaptive-RR (4.0) slightly exceeding Adaptive-Both and Fixed-None (3.8). Ultimately, Mann-Whitney U tests revealed no statistically significant differences between groups across any subjective measure.

Table 5: Questionnaire results evaluating user engagement, workload, and system usability on a 7-point Likert scale.

Scale	Group	Median	IQR
UES-SF	Fixed-None	5.4	5.40 – 5.8
	Fixed-HRV	6	5.40 – 6.6
	Adaptive-RR	5.4	4.20 – 5.4
	Adaptive-Both	5.6	5.40 – 5.8
NASA-TLX	Fixed-None	2.83	2.33 – 3.0
	Fixed-HRV	2.33	2.33 – 3.3
	Adaptive-RR	2.83	2.17 – 2.8
	Adaptive-Both	3	2.50 – 4.1
SUS	Fixed-None	3.8	3.60 – 4.0
	Fixed-HRV	4	3.70 – 4.2
	Adaptive-RR	4	3.70 – 4.4
	Adaptive-Both	3.8	3.70 – 3.9

5.4 Findings

The results demonstrated that the integration of real-time biofeedback with adaptive breathing guidance significantly improves the quality and sustainability of stress recovery. The primary finding indicates that while several guidance modes can trigger an initial relaxation response, only the fully integrated system combining both HRV biofeedback and adaptive breathing guidance allows participants to maintain a stable, elevated state of relaxation throughout the entire intervention. This suggests that the synergy of these two components is essential for converting a transient physiological peak into a prolonged period of recovery. A particularly notable outcome of this study was the system's ability to deliver these physiological benefits without significantly higher cognitive demand. Despite the increased technical complexity of the adaptive algorithms, the subjective user experience remained remarkably consistent across all experimental groups. The adaptive components did not significantly increase the perceived workload or reduce engagement and usability relative to simpler baselines. This suggests that the system successfully offloads the complexity of physiological regulation to the adaptive interface, facilitating deeper recovery without requiring additional user effort.

6 Discussion

6.1 The Significance and Versatility of Frequency Carving

The implementation of frequency carving represents a pivotal shift in how commodity earables can be transformed into sophisticated monitors without the need for specialized sensors like PPG or IMUs. By isolating a specific narrow 100 Hz band (150-250 Hz), the system creates a dedicated sensing lane within the audio stream. This effectively bypasses the traditional trade-off between high-fidelity audio playback and sensitive physiological signal acquisition.

Since frequency carving is robust enough to maintain signal accuracy without compromising perceived relaxation, it has immediate and practical applications across a wide variety of mindfulness and meditation settings. The technique's success with diverse audio types, including slow music, human voices, and nature sounds, means it can be seamlessly integrated into existing breathing exercise platforms and guided meditation apps. By removing the need for users to purchase specialized gear, frequency carving allows any standard pair of earbuds to become a tool for closed-loop biofeedback. This accessibility enables personalized, responsive breathing guidance in everyday wellness routines, transforming listening into a tool for physiological regulation.

6.2 Chord-Based Auditory Biofeedback as a Design Choice

Prior auditory biofeedback approaches in earables have largely focused on nature sounds [53, 59], discrete notes [21, 34, 62], or overlays of music and breathing sounds [34, 53]. These approaches often struggle to simultaneously convey multiple physiological signals without increasing perceptual load. Chords address this by fusing harmonically related tones into a unified auditory object [33], enabling richer encoding while preserving perceptual coherence.

Design choices within the chord framework were guided by psychoacoustic and affective principles. Major chords were selected for their consistent association with positive valence and calmness [26], and voicings were placed in the middle register where tones are perceived as warm and stable [9], which also yielded the highest user comfort ratings in preliminary testing. Beyond individual parameters, the chord representation uniquely enables dual encoding of both RR and HRV within a single acoustic signal: respiratory guidance is conveyed through a Gaussian amplitude envelope, while harmonic complexity reflects HRV progression from a single root tone to a full triad. This design leverages integer-ratio frequency relationships between chord tones to promote harmonic blending rather than competition [33], allowing added components to function as a coherent auditory reward. Together, these properties position chord-based biofeedback as the acoustic foundation of the closed-loop adaptive framework.

Importantly, chords represent one implementation choice within this framework rather than a constraint on the auditory modality. They were selected for their ability to jointly encode respiratory guidance and HRV progression within a coherent auditory stream. However, the underlying sensing and frequency carving pipeline can be extended to various other auditory forms, enabling personalized biofeedback experiences across users and contexts.

6.3 The Synergy of Cardio-Respiratory Adaptive Biofeedback

The transition from single-modality cues to a dual-factor adaptive framework addresses a critical gap in traditional biofeedback design: the inability of static or single-metric systems to sustain long-term physiological engagement. While single-factor systems typically focus on either fixed-pace breathing or HRV biofeedback in isolation, the dual-factor approach recognizes that respiration and cardiac rhythms are deeply interconnected through cardio-respiratory coupling. By modulating both breathing pace and harmonic complexity simultaneously, the system creates a more comprehensive feedback loop that mirrors the body's natural synchronization.

The necessity of this integrated approach becomes clear when examining the limitations of adaptive pacing alone. While adjusting respiratory guidance can initiate a relaxation response, the absence of a secondary reward signal (such as HRV-driven chord complexity) often leads to a gradual return to original HRV levels. By layering chord richness as an acoustic reward for improved autonomic state, the dual-factor model provides the continuous reinforcement needed to bridge the gap between a brief physiological spike and a stable, prolonged period of recovery. This suggests that breathing guidance provides the anchor, while HRV feedback provides the incentive for sustained physiological balance.

6.4 Limitations and Future Work

While Earmony provides a robust framework for adaptive biofeedback, several limitations offer avenues for future research. First, the system relies exclusively on musical chords for auditory cues. Future studies could explore alternative soundscapes, such as nature sounds or synthesized white noise for real-time modulation. Additionally, our evaluation of frequency carving focused on perceived relaxation rather than a comprehensive assessment of overall listening quality. While the results suggest that frequency carving did not significantly reduce perceived relaxation for the tested mindfulness audio, further studies are needed to examine its broader impact on factors such as audio fidelity, preference, and long-term listening experience. Another consideration involves the environmental conditions under which the system was evaluated. Like most biofeedback training, our studies were conducted in relatively quiet, controlled environments but were still subject to realistic ambient noise of around 35 dB to 55 dB. In practice, louder ambient noise or significant physical movement could potentially impact the signal-to-noise ratio of the in-ear microphone, making the detection of subtle cardiac and respiratory signals more challenging. Assessing these acoustic interference factors remains an important area for future research. Finally, integrating this closed-loop paradigm into broader digital wellness platforms such as attention regulation, performance training, and digital wellness remains a key step toward supporting physiological regulation across diverse contexts.

7 Conclusion

We presented Earmony, an adaptive auditory cardio-respiratory biofeedback system implemented on commodity earbuds that utilizes only the built-in, in-ear microphone. By introducing frequency carving, we demonstrated that it is possible to unobtrusively sense respiratory signals during active audio playback without compromising the user's perceived relaxation. The results of our user study indicate that a dual-factor adaptive approach is essential for effective physiological regulation. While single-modality systems can trigger initial relaxation, the integration of real-time HRV biofeedback with adaptive breathing guidance is necessary to maintain a sustained state of recovery after stress. Furthermore, this closed-loop intervention provides these benefits without increasing the cognitive load of the user. We envision Earmony as a step toward a future where everyday earables actively support physiological well-being, unlocking a seamless paradigm for ubiquitous health monitoring and regulation.

Acknowledgments

This work was supported by the Natural Science Foundation of China under grant No. 62472244, 62132010, the Qinghai University Research Ability Enhancement Project under grant No. 2025KTSA05, Foundation of National Key Laboratory of Human Factors Engineering under grant No. HFNKL2024W06, Undergraduate Education Innovation Grants of Tsinghua University, Tsinghua University Initiative Scientific Research Program, and Beijing Key Lab of Networked Multimedia.

References

[1] Tousif Ahmed, Md Mahbubur Rahman, Mohsin Yusuf Ahmed, Ebrahim Nemati, Minh Dinh, Nathan Folkman, Jilong Kuang, and Alex Gao. 2021. Rrmonitor: A resource-aware end-to-end system for continuous monitoring of respiration rate using earbuds. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2463–2467.

[2] Mohammed A Almazrouei, Ruth M Morgan, and Itiel E Dror. 2023. A method to induce stress in human subjects in online research environments. *Behavior research methods* 55, 5 (2023), 2575–2582.

[3] Marco Barbaresi, Davide Nardo, and Sabrina Fagioli. 2024. Physiological entrainment: a key Mind–Body mechanism for cognitive, motor and affective functioning, and Well-Being. *Brain Sciences* 15, 1 (2024), 3.

[4] Johannes Blum, Christoph Rockstroh, and Anja S Göritz. 2019. Heart rate variability biofeedback based on slow-paced breathing with immersive virtual reality nature scenery. *Frontiers in psychology* 10 (2019), 2172.

[5] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[6] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2023. heart: Motion-resilient heart rate monitoring with in-ear microphones. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 200–209.

[7] Qi Chen, Jiachen Du, Zhihao Yao, Yuanling Feng, Guanhong Liu, Ning Ni, Zhenyu Wu, Haoran Liu, Xijing Chen, and Haipeng Mi. 2025. Breath of Life: A Biofeedback Game with Diverse Breathing Techniques for Enhanced Real-Life Emotion Regulation. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.

[8] Jin-Young Choi, Seonghee Jeon, Hana Kim, Jaeyoung Ha, Gyeong-suk Jeon, Jeong Lee, and Sung-il Cho. 2022. Health-related indicators measured using earable devices: systematic review. *JMIR mHealth and uHealth* 10, 11 (2022), e36696.

[9] Marco Costa. 2013. Effects of mode, consonance, and register in visual and word-evaluation affective priming experiments. *Psychology of Music* 41, 6 (2013), 713–728.

[10] Thomas E Dick, Yee-Hsee Hsieh, Rishi R Dhingra, David M Baekey, Roberto F Galán, Erica Wehrwein, and Kendall F Morris. 2014. Cardiorespiratory coupling: common rhythms in cardiac, sympathetic, and respiratory activities. *Progress in brain research* 209 (2014), 191–205.

[11] Abbey L Dvorak and Eugenia Hernandez-Ruiz. 2021. Comparison of music stimuli to support mindfulness meditation. *Psychology of Music* 49, 3 (2021), 498–512.

[12] 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.

[13] Andrea Ferlini, Alessandro Montanari, Chulhong Min, Hongwei Li, Ugo Sassi, and Fahim Kawsar. 2021. In-ear PPG for vital signs. *IEEE Pervasive Computing* 21, 1 (2021), 65–74.

[14] Dana L Frank, Lamees Khorshid, Jerome F Kiffer, Christine S Moravec, and Michael G McKee. 2010. Biofeedback in medicine: who, when, why and how? *Mental health in family medicine* 7, 2 (2010), 85.

[15] Alfredo J Garcia III, Jenna E Koschnitzky, Tatiana Dashevskiy, and Jan-Marino Ramirez. 2013. Cardiorespiratory coupling in health and disease. *Autonomic Neuroscience* 175, 1-2 (2013), 26–37.

[16] Jean-Marie Grégoire, Cédric Gilon, Stéphane Carlier, and Hugues Bersini. 2023. Autonomic nervous system assessment using heart rate variability. *Acta cardiologica* 78, 6 (2023), 648–662.

[17] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[18] Changshuo Hu, Thivya Kandappu, Yang Liu, Cecilia Mascolo, and Dong Ma. 2024. BreathPro: Monitoring breathing mode during running with earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–25.

[19] 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).

[20] Hugh Hunkin, Daniel L King, and Ian T Zajac. 2021. EEG neurofeedback during focused attention meditation: Effects on state mindfulness and meditation experiences. *Mindfulness* 12, 4 (2021), 841–851.

[21] Ennio Idrobo-Avila, Humberto Loaiza-Correa, Flavio Munoz-Bolanos, Leon van Noorden, and Rubiel Vargas-Canas. 2022. Development of a biofeedback system using harmonic musical intervals to control heart rate variability with a generative adversarial network. *Biomedical Signal Processing and Control* 71 (2022), 103095.

[22] Muammar M Kabir, Hany Dimitri, Prashanthan Sanders, Ral Antic, Eugene Nalivaiko, Derek Abbott, and Mathias Baumert. 2010. Cardiorespiratory phase-coupling is reduced in patients with obstructive sleep apnea. *PLoS One* 5, 5 (2010), e10602.

[23] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation* 15, 3 (2018), 235.

[24] Bonhee Ku, Tatsuya Itagaki, and Katie Seaborn. 2023. Dis/immersion in mindfulness meditation with a wandering voice assistant. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.

[25] Michael Küttner, Valeria Zitz, Supraja Ramesh, Michael Beigl, and Tobias Röddiger. 2026. EarResp-ANS: Audio-Based On-Device Respiration Rate Estimation on Earphones with Adaptive Noise Suppression. *arXiv preprint arXiv:2602.03549* (2026).

[26] Carolina Labbé, Wiebke Trost, and Didier Grandjean. 2021. Affective experiences to chords are modulated by mode, meter, tempo, and subjective entrainment. *Psychology of Music* 49, 4 (2021), 915–930.

[27] Paul M Lehrer and Richard Gevirtz. 2014. Heart rate variability biofeedback: how and why does it work? *Frontiers in psychology* 5 (2014), 756.

[28] 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.

[29] Rong-Hao Liang, Bin Yu, Mengru Xue, Jun Hu, and Loe MG Feijs. 2018. BioFidget: Biofeedback for respiration training using an augmented fidget spinner. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[30] Stefan Liebich and Peter Vary. 2021. Occlusion effect cancellation in headphones and hearing devices—The sister of active noise cancellation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2021), 35–48.

[31] Yang Liu, Kayla-Jade Butkow, Jake Stuchbury-Wass, Adam Pullin, Dong Ma, and Cecilia Mascolo. 2025. Respear: Earable-based robust respiratory rate monitoring. In *2025 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 67–77.

[32] Yang Liu, Qiang Yang, Kayla-Jade Butkow, Jake Stuchbury-Wass, Dong Ma, and Cecilia Mascolo. 2025. EarMeter: Continuous Respiration Volume Monitoring with Earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 4 (2025), 1–29.

[33] Pieter-Jan Maes, Jeska Buhmann, and Marc Leman. 2016. 3Mo: a model for music-based biofeedback. *Frontiers in neuroscience* 10 (2016), 548.

[34] Georgios Marentakis, Debanjan Borthakur, Paul Batchelor, Judith P Andersen, and Victoria Grace. 2021. Using breath-like cues for guided breathing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.

[35] Michela Masè, Alessandro Micarelli, and Giacomo Strapazzon. 2020. Hearables: new perspectives and pitfalls of in-ear devices for physiological monitoring. A scoping review. *Frontiers in Physiology* 11 (2020), 568886.

[36] Malahat HK Mehrban, Jérémie Voix, and Rachel E Bouserhal. 2024. Classification of breathing phase and path with in-ear microphones. *Sensors* 24, 20 (2024), 6679.

[37] Abele Michela, Jacobien M van Peer, Jan C Brammer, Anique Nies, Marieke MJW van Rooij, Robert Oostenveld, Wendy Dorrestijn, Annika S Smit, Karin Roelofs, Floris Klumpers, et al. 2022. Deep-breathing biofeedback trainability in a virtual-reality action game: a single-case design study with police trainers. *Frontiers in Psychology* 13 (2022), 806163.

[38] Margaret Morris and Farzin Guilak. 2009. Mobile heart health: project highlight. *IEEE Pervasive Computing* 8, 2 (2009), 57–61.

[39] Heather L O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.

[40] Nikola N Radovanović, Siniša U Pavlović, Goran Milašinović, Bratislav Kirćanski, and Mirjana M Platiša. 2018. Bidirectional cardio-respiratory interactions in heart failure. *Frontiers in physiology* 9 (2018), 165.

[41] Md Mahbubur Rahman, Tousif Ahmed, Mohsin Yusuf Ahmed, Minh Dinh, Ebrahim Nemati, Jilong Kuang, and Jun Alex Gao. 2022. Breathebuddy: Tracking real-time breathing exercises for automated biofeedback using commodity earbuds. *Proceedings of the ACM on human-computer interaction* 6, MHCI (2022), 1–18.

[42] Md Mahbubur Rahman, Tousif Ahmed, Mohsin Yusuf Ahmed, Ebrahim Nemati, Minh Dinh, Nathan Folkman, Md Mehedi Hasan, Jilong Kuang, and Jun Alex Gao. 2021. Towards motion-aware passive resting respiratory rate monitoring using earbuds. In *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 1–4.

[43] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. 2022. Sensing with earables: A systematic literature review and taxonomy of phenomena. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 6, 3 (2022), 1–57.

[44] Tobias Röddiger, Daniel Wolffram, David Laubenstein, Matthias Budde, and Michael Beigl. 2019. Towards respiration rate monitoring using an in-ear headphone inertial measurement unit. In *Proceedings of the 1st International Workshop on Earable Computing*. 48–53.

[45] Julia Romero, Andrea Ferlini, Dimitris Spathis, Ting Dang, Katayoun Farrahi, Fahim Kawsar, and Alessandro Montanari. 2024. OptiBreathe: An earable-based PPG system for continuous respiration rate, breathing phase, and tidal volume monitoring. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*. 99–106.

[46] Ramesh Kumar Sah, Md Mahbubur Rahman, Viswam Nathan, Li Zhu, Jungmok Bae, Christina Rosa, Wendy Berry Mendes, Jilong Kuang, and Alex Jun Gao. 2024. Heart rate variability estimation with dynamic fine filtering and global-local context outlier removal. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2091–2095.

[47] Mark S Schwartz and Frank Andrasik. 2017. *Biofeedback: A practitioner's guide*. Guilford Publications.

[48] Chen-Hsuan Shih, Naofumi Tomita, Yanick X Lukic, Álvaro Hernández Reguera, Elgar Fleisch, and Tobias Kowatsch. 2019. Breeze: Smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breathing training. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 4 (2019), 1–30.

[49] Patrick R Steffen, Tara Austin, Andrea DeBarros, and Tracy Brown. 2017. The impact of resonance frequency breathing on measures of heart rate variability, blood pressure, and mood. *Frontiers in public health* 5 (2017), 222.

[50] Kazuhiro Taniguchi and Atsushi Nishikawa. 2018. Earable POCER: Development of a point-of-care ear sensor for respiratory rate measurement. *Sensors* 18, 9 (2018), 3020.

[51] Grace Theodore, Jing-Yuan Huang, Lung-Pan Cheng, and Yi-Ping Hung. 2024. Breathm: a calm device with personalized slow breathing guidance. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–5.

[52] Yvonne M Ulrich-Lai and James P Herman. 2009. Neural regulation of endocrine and autonomic stress responses. *Nature reviews neuroscience* 10, 6 (2009), 397–409.

[53] Bavo Van Kerrebroeck and Pieter-Jan Maes. 2021. A breathing sonification system to reduce stress during the COVID-19 pandemic. *Frontiers in Psychology* 12 (2021), 623110.

[54] Evgeny G Vaschillo, Bronya Vaschillo, and Paul M Lehrer. 2006. Characteristics of resonance in heart rate variability stimulated by biofeedback. *Applied psychophysiology and biofeedback* 31, 2 (2006), 129–142.

[55] Lin Xie, Mengjun Li, Shijie Dang, Chaomin Li, Xiaoni Wang, Binbin Liu, Mengqi Mei, and Jianbao Zhang. 2018. Impaired cardiorespiratory coupling in young normotensives with a family history of hypertension. *Journal of Hypertension* 36, 11 (2018), 2157–2167.

[56] Yuchen Xu, Abhinav Uppal, Min Suk Lee, Kuldeep Mahato, Brian L Wuerstle, Muyang Lin, Omeed Djassemi, Tao Chen, Rui Lin, Akshay Paul, et al. 2024. Earable multimodal sensing and stimulation: a prospective toward unobtrusive closed-loop biofeedback. *IEEE Reviews in Biomedical Engineering* 18 (2024), 5–25.

[57] Bin Yu, Pengcheng An, Sjoerd Hendriks, Ning Zhang, Loe Feijs, Min Li, and Jun Hu. 2021. ViBreathe: Heart rate variability enhanced respiration training for workaday stress management via an eyes-free tangible interface. *International Journal of Human–Computer Interaction* 37, 16 (2021), 1551–1570.

[58] Bin Yu, Loe MG Feijs, Mathias Funk, and Jun Hu. 2015. Designing Auditory Display of Heart Rate Variability in Biofeedback Context.. In *ICAD*. 294–298.

[59] Bin Yu, Mathias Funk, Jun Hu, and Loe Feijs. 2018. Unwind: a musical biofeedback for relaxation assistance. *Behaviour & Information Technology* 37, 8 (2018), 800–814.

[60] Ruotong Yu, Xintong Wu, Lily Sheng, Zihao Dong, Yuntao Wang, and Yuanchun Shi. 2026. Enabling Adaptive Cardio-Respiratory Biofeedback Training on Ubiquitous Hand-Worn Devices. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. 1–18.

[61] Julien Zanin, Dani Tomlin, and Gary Rance. 2024. Effectiveness of Noise Cancelling Earbuds in Reducing Hearing and Auditory Attention Deficits in Children with Autism. *Journal of Clinical Medicine* 13, 16 (2024), 4786.

[62] Sebastian Zepf, Neska El Haouij, Jinmo Lee, Asma Ghandeharioun, Javier Hernandez, and Rosalind W Picard. 2020. Studying personalized just-in-time auditory breathing guides and potential safety implications during simulated driving. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*. 275–283.

[63] Zixuan Zhang, Wenxuan Jin, Dejiao Huang, and Zhongwei Sun. 2025. A Novel Audio-Perception-Based Algorithm for Physiological Monitoring. *Sensors* 25, 12 (2025), 3582.

[64] Yuzheng Zhu, Zhangxin Liang, Jie Zheng, Yongpan Zou, Victor CM Leung, and Kaishun Wu. 2025. VitalEar: an Earable Heartbeat and Respiratory Rate Monitoring System under Aerobic Exercises. *IEEE Transactions on Mobile Computing* (2025).

A Use of Generative AI Tools

For this work, generative AI tools were used in a limited and responsible manner:

- Text: ChatGPT 5.6 and Gemini 3.6 were used only for grammar checking and language polishing to improve the clarity and the readability of the manuscript.
- Figures: We used ChatGPT 5.6 for generating part of the illustrative teaser image in Figure 1 to visually support the manuscript.

No AI-generated content was used to generate experimental data, results, or any interpretive content.