

APG: Audioplethysmography for Cardiac Monitoring in Hearables

Xiaoran Fan
Google

David Pearl
Google

Longfei Shangguan
University of Pittsburgh

Richard Howard
Rutgers University

Trausti Thormundsson
Google

ABSTRACT

This paper presents Audioplethysmography (APG), a novel cardiac monitoring modality for active noise cancellation (ANC) headphones. APG sends a low intensity ultrasound probing signal using an ANC headphone's speakers and receives the echoes via the on-board feedback microphones. We observed that, as the volume of ear canals slightly changes with blood vessel deformations, the heartbeats will modulate these ultrasound echoes. We built mathematical models to analyze the underlying physics and propose a multi-tone APG signal processing pipeline to derive the heart rate and heart rate variability in both constrained and unconstrained settings. APG enables robust monitoring of cardiac activities using mass-market ANC headphones in the presence of music playback and body motion such as running.

We conducted an eight-month field study with 153 participants to evaluate APG in various conditions. Our studies conform to the (Institutional Review Board) IRB policies from our company. The presented technology, experimental design, and results have been reviewed and further improved by feedback garnered from our internal Health Team, Product Team, User Experience (UX) Team and Legal team. Our results demonstrate that APG achieves consistently high HR (3.21% median error across 153 participants in all scenarios) and HRV (2.70% median error in interbeat interval, IBI) measurement accuracy. Our UX study further shows that APG is resilient to variation in: skin tone, sub-optimal seal conditions, and ear canal size.

CCS CONCEPTS

- Human-centered computing → Mobile computing;
- Hardware → Emerging interfaces.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ACM MobiCom '23, October 2–6, 2023, Madrid, Spain

© 2023 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9990-6/23/10...\$15.00

<https://doi.org/10.1145/3570361.3613281>

KEYWORDS

Earable Computing; Heart Rate Monitoring; Wearable Devices; Mobile Health

ACM Reference Format:

Xiaoran Fan, David Pearl, Longfei Shangguan, Richard Howard, and Trausti Thormundsson. 2023. APG: Audioplethysmography for Cardiac Monitoring in Hearables. In *The 29th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '23), October 2–6, 2023, Madrid, Spain*. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3570361.3613281>

1 INTRODUCTION

Hearables are embedded devices that can be placed in or around the ear to unobtrusively monitor human health and activities [27, 47]. To date, hearables as a health-sensing system is mostly realized by packing a plethora of sensors (e.g., Photoplethysmography (PPG) and Electromyography (EMG) [3, 8, 10, 11, 37]) with a micro-controller to enable health applications such as sleep monitoring [46], heart rate and blood pressure tracking [24]. However, such a *sensor mounting paradigm* inevitably adds cost, weight, power consumption, acoustic design complexity, and form factor to hearables, constituting a strong barrier to its wide adoption.

Recently, however, the rapid growth of active noise cancellation (ANC) headphones may lead to innovation in the mobile health-sensing field. ANC headphones put a feedback microphone inside the ear cup to sense the environment noise heard by the user. As this microphone hears the noise similar to the person does, the ANC circuit can create the anti-noise before sending the resulting signal to the headset speaker. To improve noise cancellation, the ANC headphones further leverage a feedforward microphone outside the ear cup to work collaboratively with the feedback microphone to expand the ANC bandwidth. The microphones open up new opportunities for many sensing applications. By taking advantage of feedforward and feedback microphone on ANC headphones, it is feasible to detect or record many bio-signals inside the ear canal (e.g., heart sound). This *passive sensing paradigm* has been extensively exploited in academia, prompting many exciting mobile applications including heart rate monitoring, ear diseases diagnosis, respiration monitoring, and body activity recognition [25, 36, 39, 58, 59].

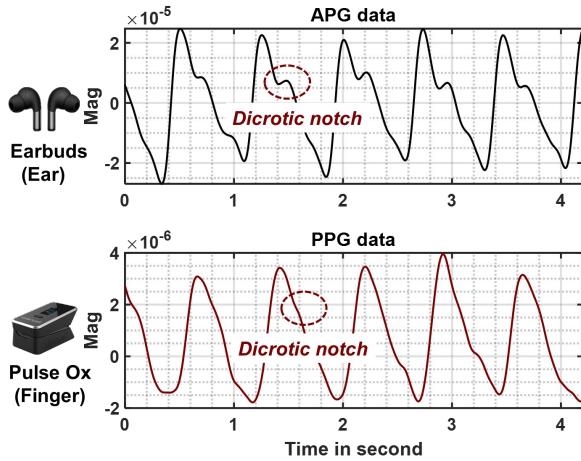


Figure 1: APG signals vs. PPG signals. APG waveform displays similarly as the PPG waveform. APG captures high fidelity cardiac activities such as the dicrotic notch – a distinctive feature of the pressure waveform in the central arteries.

Sensing limit in ANC headphones – from manufacturers’ perspectives. The microphones on consumer-grade ANC headphones are tightly integrated with the proprietary ANC chips, which prohibits mobile users from accessing their readings unless they break the ear cup (or earbud) and wire out the signals using an external audio line. To circumvent this constraint, prior works [32, 33, 39, 42, 43, 45, 54, 58] put a dedicated microphone inside the ear canal to sense physiological signals passively (e.g., recording the low-frequency heartbeats or body motions). Unfortunately, the hearable features and applications demonstrated in these systems are challenging to be deployed on commercial ANC headphones due to a plethora of practical constraints, as elaborated below.

Firstly, to ensure ANC operation, the microphones in consumer-grade ANC headphones come with built-in analog (preferred) or digital high-pass filters to prevent them from being saturated by the body motions or strong wind noise (can be louder than 120 dB SPL) [31]. Saturated microphones will reduce the microphone dynamic range and create low-frequency boomy sound, which compromising the ANC operation. More importantly, compromised ANC might cause **howling** [35] on the speaker, which might lead to hearing damage. As such, many hearables features that rely on low-frequency signals (≤ 50 Hz) are challenging to work with commercial ANC headphones. These features and applications include most physiology sensing signals such as heart rate sensing. Secondly, most consumer-grade ANC headphones are designed in a true wireless stereo (TWS) manner. They have independent chips on the left channel and right channel that process acoustic signals independently. There is no centralized controller that processes dual-channel acoustic signals together. Accordingly, some advanced research

that require two earbuds [23, 26] are challenging to deploy commercial products. Last but not least, passive hearable sensing suffers severely from music interference (including nonlinear interference that is hard to remove). Therefore, these hearable features and applications are challenging to work robustly in the presence of music playback.

This paper presents the design and implementation of APG, a novel acoustic sensing system that enables cardiac activity monitoring using commodity ANC headphones under aforementioned constraints. The basic idea is to manipulate the speakers of ANC headphones to *actively* emit a set of low-intensity probing signals at different ultrasonic frequencies. These probing signals work above the human audible band but can be well captured by the feedback microphone even in the presence of music playback and built-in high-pass microphone filters. The key insight is that the signal received by the feedback microphone is the superposition of multipath signals reflected by the ear canal, and each path has a time-varying time delay (phase) and time-varying amplitude due to the minute dynamics of blood vessel deformation from cardiac activities. This correlation leaves an opportunity to detect cardiac activities through signal variation analysis.

However, realizing the above idea faces multiple design and implementation challenges, as elaborated below.

(i) Firstly, blood vessel deformations due to cardiac activities are usually imperceptible; they only modulate a very tiny part of the probing signal. Accordingly, the cardiac signal is invisible if we directly look at the time/frequency domain of the ultrasound receptions.

(ii) Secondly, we should send the ultrasound in low intensity (preferably much lower than the safety limit that is specified by INIRC, ACGIH, and IRPA [29, 40, 44]) to ensure the long-term safe operation of APG and enough headroom for audible sound (e.g., music). This safety constraint significantly brings down the signal-to-noise ratio (SNR) for APG, setting a strong barrier to cardiac activity detection.

(iii) Thirdly, APG should be resilient to different earbuds seal conditions as the way the earbuds seal with human ears differs from users due to the ear structure heterogeneity. Even for the same user, each time the seal is likely to change when the user wears the earbuds. The seal condition could affect the quality of the ultrasound signal drastically, which in turn impairs the robustness of cardiac sensing.

(iv) Fourthly, the interference signals such as body motions artifacts also will propagate through the human body, arriving at the ear canal. These interference signals will be modulated by the APG probing signals, overwhelming the weak cardiac signal. It is thus crucial to resolve the weak cardiac signal from signals with strong interference.

To address these challenges, we take advantage of coherent detection from the wireless communications field to mitigate noises and extract the minute cardiac activity.

The coherent detection operation allows us to extract subtle cardiac signals that are not visible in the raw ultrasound receptions. Moreover, such a design also enables us to measure cardiac activities in the presence of music playback, making long-term cardiac monitoring on hearables feasible. We then propose a multi-tone APG, an enhanced probing signal design that explores the rich ultrasound bandwidth resources to enhance the robustness and accuracy of cardiac signal detection. To enable heart rate monitoring in an unconstrained environment (e.g., head moving, running), we further leverage the frequency diversity of this multi-tone APG to separate the weak cardiac signal from the overwhelming body motion artifacts through blind source separation using just one earbud. Lastly, although APG measures the cardiac activities in ultrasound, we downsample the waveform to 100 Hz to be compatible with Bluetooth.

Further scrutiny of the cardiac signal shows that the cardiac activities modulate less than 1% of the ultrasound probing signal, which is strikingly similar to PPG¹. The fact that cardiac activities modulate less than 1% of the ultrasound probing signal is not surprising, given under the hood APG and PPG both measure changes in the volume of blood vessels in different parts of the body. The similarity of cardiac waveforms reported by APG and a FDA-approved PPG sensor also confirms our analysis (Figure 1). To examine the safety of APG, we conducted extensive safety compliance studies using the International Electrotechnical Commission (IEC) complied G.R.A.S RA0404 high-frequency ear simulator [5] with an Audio Precision APx585 analyzer [15]. Results show that APG can detect heart rate and heart rate variability (HRV)² at a high accuracy when the sound intensity measured at the ear-drum reference point (DRP) is as low as 30 dB SPL – 80 dB lower than the safety limit (110 dB).

In summary, this paper makes the following contributions:

- We present the design and implementation of a novel wearable sensing system, APG, for cardiac activities monitoring under the design constraints of mass-market TWS ANC headphones. APG can reliably detect the heart rate during music playback. APG transforms any TWS ANC headphones into smart sensing headphones with a software upgrade.
- We propose a plethora of signal processing algorithms to extract the minute cardiac signals in extremely low SNR conditions. By exploiting the frequency diversity, we further design a signal conditioning pipeline utilizing a blind source separation algorithm to resolve the cardiac activity signal under strong body interference(s).
- We tested APG on both commercially available ANC earbuds and our own experimental prototypes. We conducted

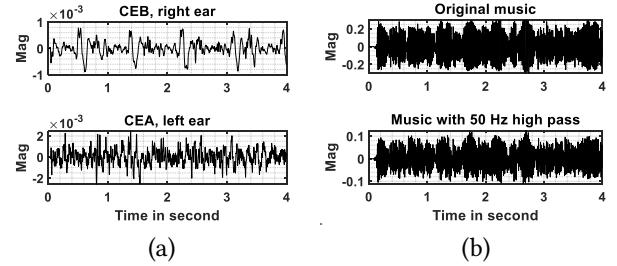


Figure 2: The limitation of passive sensing using feedback microphones on commodity ANC headphones. (a) The microphone readings wired out from two different ANC earphones in the absence of music playback. (b) The cardiac activities are overwhelmed by the music signal.

rigorous experiments designed by our UX research team on 153 users. APG achieves 3.21% median error for HR monitoring for all 153 subjects across all scenarios and 2.70% median error for HRV monitoring under various experiment settings. APG works when the music is playing and the transmitted ultrasound intensity is 80 dB lower than the aforementioned safety limit. We also observed that APG works robustly with bad earbud sealing. APG consumes negligible extra power as the APG signal is very low intensity and the audio chain is always on during noise canceling or transparency mode.

2 THE LIMITS OF PASSIVE SENSING

As mentioned briefly, the efficacy of passive sensing based on microphones is largely constrained on commodity ANC headphones. In this section, we explain these practical constraints with comprehensive benchmark studies. Specifically, we first demonstrate that not all signals wired out from commodity ANC headphones contain low-frequency cardiac activity signals. We then demonstrate that the music playback will obstruct the detection of cardiac signals. In this section, we apply a low-pass filter with 25 Hz cutoff frequency to all received signals to better illustrate the cardiac patterns.

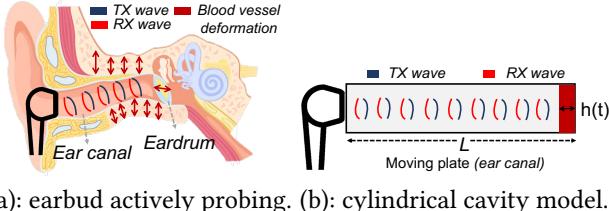
Experiment setups. We wire out the output signal of the feedback microphone on a pair of ANC headphones – commercial earbuds A (CEA)³. Likewise, we also wire out the output signal of the feedback microphone on a commercial earbuds B (CEB) for comparison. To ensure the active noise canceling (avoiding saturated microphones), the CEA buds have an analog high-pass filter built into its feedback microphone whereas the CEB earbud filters the readings of the feedback microphone in its DSP. Therefore, we expect to see heart pulses on the microphone readings wired out from CEB earbuds but not from the CEA buds.

Experiment results. A volunteer is asked to wear a CEA on their left ear and a CEB on their right ear. We then plot

¹ The cardiac-synchronized AC signal in PPG is less than 1% of DC level [9].

² HRV measurement has a higher requirement on the quality of cardiac waveform than heart rate measurement (§4.1).

³ We retracted the name of all commercial earbuds in our paper due to the company paper publishing policy.



(a): earbud actively probing. (b): cylindrical cavity model.

Figure 3: Illustrations of (a) ANC earbuds actively probe the ear canal and (b) a simplified model to describe the physiology activities inside the ear canal.

microphone readings wired out from their feedback microphones in the absence of music playback. The results are shown in Figure 2(a). As expected, we can see clear heart pulses on CEB's microphone readings. In contrast, there is no visible heart pulse in CEA's microphone readings.

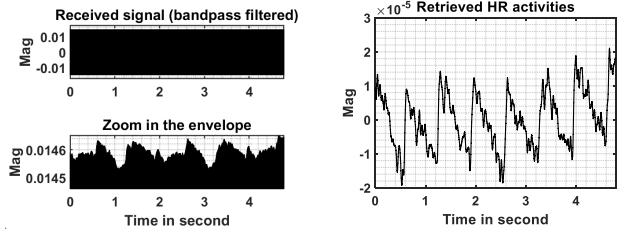
We further plot the signals captured by CEB earbuds in the presence of music playback. As shown in the top figure of Figure 2(b), the received acoustic signal is dominated by the low-frequency components from the music. The cardiac signals, accordingly, are fully overwhelmed by this strong music interference. We then play a piece of music with no acoustic component below 50Hz on the CEB earbuds and draw the signal received by the CEB earbuds' feedback microphone. The goal is to check whether we can successfully receive the cardiac pattern in the presence of non-overlapping music interference. The bottom figure of Figure 2(b) shows the result. Unfortunately, we still cannot observe the heartbeat from the received signals. This phenomenon is due to the movement of the non-linear voice coil inside the headphone driver [38]. Comparing the Y-axes of Figure 2(a) with Figure 2(b), we can see the low-frequency signal in the presence of music playback is over 100 times stronger than the cardiac signal. Accordingly, the weak cardiac signals are deeply overwhelmed by the music interference.

3 APG FOR PHYSIOLOGICAL SENSING

APG is an *active probing-based* acoustic sensing system that can work with commodity ANC headphones under aforementioned hardware constraints. In this section, we first build a physical model to explain the APG sensing (§3.1). We then present the signal detection algorithms (§3.2), and an in-depth analysis of the proposed solution (§3.3).

3.1 APG Physical Model

Physiological principle. Figure 3(a) demonstrates how an ANC earbud sends acoustic signals and receives the echoes in the ear canal. Both eardrum and tissues surrounding the ear canal will *slightly squeeze* the ear canal cavity due to the blood vessel deformation. This volumetric change in the ear canal inherently modulates the acoustic probing signal in the ear canal. Note that the blood vessels deforms according



(a): the raw echo receptions. (b): retrieved heart rate signals.

Figure 4: (a) The 30 kHz band-pass in-ear echo signal received by the feedback microphone. The cardiac activities are undetectable in this raw signal. (b) The cardiac pattern is retrieved after coherent detection.

to the cardiac activities [51]. Therefore, we can retrieve the heart rate pattern by examining the echo signals.

Signal propagation model. To understand how blood vessel deformation modulates echoes, we build a cylindrical cavity model to analyze this process. As shown in Figure 3(b), we model the ear canal using a cylindrical tube with a fixed radius and one open end. The speaker and microphone are placed at this open end while the eardrum closes this tube at the other end. The closed end (*i.e.*, the red plate shown in Fig 3 (b)) moves back and forth to emulate the volumetric change of ear canal due to blood vessel deformation.

We define the acoustic wave sent from the earphone speaker as $S_t(t) = \cos(\Omega_{fc}t + \Phi_{ini})$, where Ω_{fc} is the signal frequency; Φ_{ini} is the initial phase of this probing signal. The echo $S_e(t)$ can be written as⁴:

$$S_e(t) = \alpha \cos\left(\Phi_{ini} + \Phi_r + \Omega_{fc}\left(t + \frac{2L+h}{v}\right)\right). \quad (1)$$

where L is the length of the tube; h is the displacement of the plate; v is the speed of sound. Φ_r is the phase change when the signal bounces off the ear canal. α is the amplitude attenuation. The received signal $S_r(t)$ at the feedback microphone is the superposition of $S_e(t)$ and $S_t(t)$, which can be represented as:

$$S_r(t) = S_e(t) + S_t(t) \\ = N + \cos(\phi_{ini} + \Omega_{fc} t) + \alpha \cos\left(\phi_r + \phi_{ini} + \Omega_{fc}\left(t + \frac{2L+h}{v}\right)\right) \quad (2)$$

where N is the noise term. From the above equation we can see that the displacement h of the plate due to cardiac activities modulates the received signal $S_r(t)$, demonstrating the feasibility of retrieving cardiac activities from the received acoustic signal. However, in practice, as the volumetric change of the ear canal due to cardiac activities is very subtle (*i.e.*, $h \ll L$), the cardiac modulation would be extremely weak and thus cannot be directly observed from the original time domain signal, as shown in Figure 4(a).

⁴ For ease of analysis, we assume there is only one echo signal.

3.2 APG Mixing

To retrieve weak cardiac modulation signals in low SNR conditions, we apply coherent detection by multiplying the received signal $S_r(t)$ with the reference signal. Here the reference signal is the probing signal generated locally. Let the in-phase part of the reference signal be $S_I(t) = \cos(\Omega_{fc} t)$. The quadrature-phase part can be derived similarly: $S_Q(t) = \cos(\frac{\pi}{2} + \Omega_{fc} t)$. The mixing process can be written as:

$$\begin{aligned} S_{APG}(t)S_I(t) &= \frac{\cos(\phi_{ini} + 2\Omega_{fc} t)}{2} + \frac{\cos(\phi_{ini})}{2} + N \cos(\Omega_{fc} t) \\ &\quad + \frac{\alpha \cos(\phi_r + \phi_{ini} + \frac{2L\Omega_{fc}}{v} + \frac{\Omega_{fc} h}{v})}{2} \\ &\quad + \frac{\alpha \cos(\phi_r + \phi_{ini} + 2\Omega_{fc} t + \frac{2L\Omega_{fc}}{v} + \frac{\Omega_{fc} h}{v})}{2}. \end{aligned} \quad (3)$$

Note Ω_{fc} is the frequency of the ultrasound probing signal ($f_{fc} \geq 20\text{kHz}$). We next filter the mixed signal $S_{APG}(t)S_I(t)$ using a low pass filter with low cutoff frequency (20Hz by default). The high frequency terms in Eqn 3, $\frac{\cos(\phi_{ini}+2\Omega_{fc} t)}{2}$ and $\frac{\alpha \cos(\phi_r+\phi_{ini}+2\Omega_{fc} t+\frac{2L\Omega_{fc}}{v}+\frac{\Omega_{fc} h}{v})}{2}$, will be inherently removed. The noise term $N \cos(\Omega_{fc} t)$ will contains only in-band noise part, namely, N_{IBI} and N_{IBQ} . The in-phase and quadrature-phase part of the mixed-down signal I at the baseband are represented by:

$$\begin{cases} I &= \frac{\cos(\phi_{ini})}{2} + \frac{\alpha \cos(\frac{\Omega_{fc} h + \phi_r v + \phi_{ini} v + 2L\Omega_{fc}}{v})}{2} + N_{IBI} \\ Q &= \frac{\sin(\phi_{ini})}{2} + \frac{\alpha \sin(\frac{\Omega_{fc} h + \phi_r v + \phi_{ini} v + 2L\Omega_{fc}}{v})}{2} + N_{IBQ} \end{cases} \quad (4)$$

Based on Euler's formula, the amplitude R and phase Φ of the mixed-down APG signal can be represented by:

$$R = \sqrt{I^2 + Q^2} = \frac{\sqrt{\alpha^2 + 2 \cos\left(\frac{\Omega_{fc} h + \phi_r v + \phi_{ini} v + 2L\Omega_{fc}}{v}\right) \alpha + 1}}{2} + N_{Ares} \quad (5)$$

$$\begin{aligned} \Phi &= \arctan \frac{Q}{I} \\ &= \arctan \left(\frac{\sin(\phi_{ini}) + \alpha \sin\left(\frac{\Omega_{fc} h + \phi_r v + \phi_{ini} v + 2L\Omega_{fc}}{v}\right)}{\cos(\phi_{ini}) + \alpha \cos\left(\frac{\Omega_{fc} h + \phi_r v + \phi_{ini} v + 2L\Omega_{fc}}{v}\right)} \right) + N_{Pres} \end{aligned} \quad (6)$$

In the above equation, N_{Ares} and N_{Pres} are the residue noise in the amplitude and phase of the mixed-down signal, respectively. From Equation 5 and Equation 6, we have two observations. First, the amplitude R and phase Φ of the mixed-down signal change with the cardiac amplitude modulation (denoted by h in the equation). Second, the original noise term N is reduced to in-band residual noise N_{res} after the

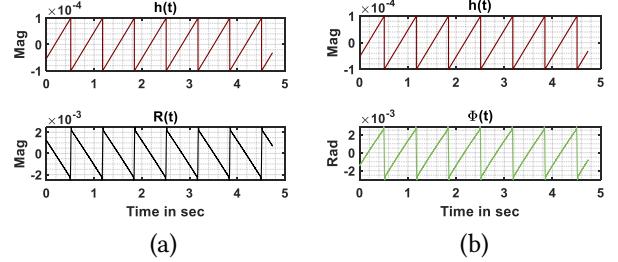


Figure 5: Simulation based on the APG physical model. The cardiac activities modulate both the phase and amplitude of the mixed-down signal. (a): The simulated cardiac stimulus function $h(t)$ vs. the amplitude of the mixed-down signals $R(t)$. (b): The simulated cardiac stimulus function $h(t)$ vs. the phase of the mixed-down signals $\Phi(t)$.

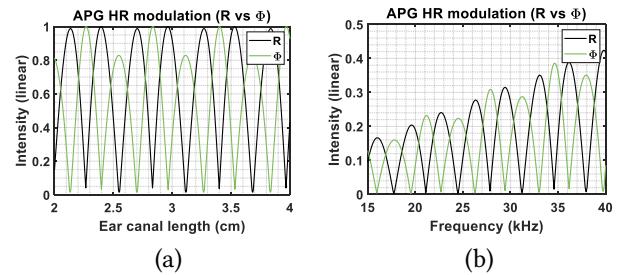


Figure 6: The impact of the length of the ear canal and the frequency of probing signals on the cardiac modulation. The phase modulation and amplitude modulation show up alternatively and they never align at the same point. (a): The cardiac modulation vs ear canal length. (b): The cardiac modulation vs the APG frequency.

coherent detection (comparing Equation 3 with Equation 4), which brings SNR gain to the cardiac modulation signal.

Model validation. To validate the efficacy of the above operation, we conduct a real-world experiment by sending a 30 kHz ultrasound signal into the ear canal using a pair of CEA earbuds. We then leverage the feedback microphone on this CEA Buds to receive the echoes and wire out the readings. Figure 4 shows the echoes before and after the coherent detection (i.e., signal mixing). We observe a clear heartbeat pattern in the mixed-down signal, whereas the signal before coherent detection is dominated by noises.

3.3 An In-Depth Analysis of APG Sensing

To understand how the phase Φ and amplitude R of the mixed-down signals change with the cardiac amplitude modulations, we use a sawtooth function [13] to simulate the cardiac stimulus function $h(t)$ and plot the phase and amplitude of the mixed-down signal as a function of time, ear canal length, and the carrier frequency of the probing signals. The simulation results are shown in Figure 5 and Figure 6. From these results we have three key observations.

(i): Firstly, in accordance with our analysis, the cardiac activities $h(t)$ modulate both phase and amplitude of the mixed-down signals $\Phi(t)$ and $R(t)$ respectively, as shown in Figure 5. These phase and amplitude profiles manifest a clear repetitive pattern, with each repetition corresponding to a heartbeat (these signals can be phase inverted).

(ii): Secondly, as shown in Figure 6(a), the intensity of both phase modulation and amplitude modulation changes with the length of the ear canal (*i.e.*, different ear structure). It is challenging to detect heartbeat cycles in the presence of low phase or amplitude modulation intensity. However, the intensity dips on phase modulation and amplitude modulation appear alternatively and they never align at the same ear canal length setting. Accordingly, given an ear canal with an arbitrary length (structure), we can always detect cardiac activities either in the phase modulation profile or in the amplitude modulation profile. This observation also indicates that APG works robustly under various earbuds seal conditions because a different seal translates to a change in the equivalent ear canal length.

(iii): Thirdly, from Figure 6(b) we observe that both phase modulation and amplitude modulation intensify as we increase the frequency of probing signals. This is because a higher frequency translates to a shorter wavelength and a shorter wavelength is more sensitive to the minute change of ear canal structure. Moreover, similar to the observation (ii), we observe a similar trend of the phase modulation and amplitude modulation alternation over frequencies.

We also observed this alternative amplitude and phase modulation behavior in real-world experiments. For instance, Figure 7 shows the amplitude and phase profile of two mixed-down signal receptions, respectively. These two signals are collected in two separate experiments. In accordance with our simulation, in the first experiment (Figure 7(a)), the cardiac modulation is only observable on the amplitude profile of received signals. While in the second experiment (Figure 7(b)) the phase profile manifests a strong repetitive pattern whereas the amplitude profile is not.

4 APG SENSING IN PRACTICE

The above section explains the rationale of APG on cardiac activity sensing. In this section, we address application-specific challenges in both constrained (§4.1) and unconstrained environments (§4.2), respectively.

4.1 Heart Rate Variability Measurement

Heart Rate Variability (HRV) refers to the frequency of interbeat intervals (RR intervals) changes over time [56]. HRV are critical biomarkers that can reveal heart conditions and mental health issues such as anxiety and depression [50, 53]. The standard way to measure HRV is to calculate the Root Mean Square of Successive Differences (RMSSD) between

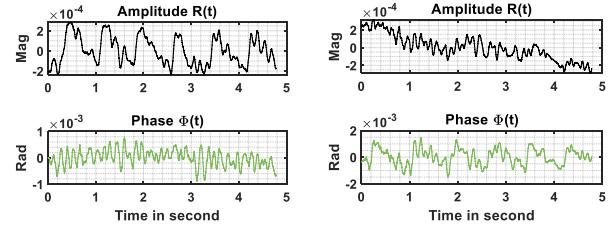


Figure 7: APG measurement results from a 30 kHz in-ear ultrasound after IQ mixing. The cardiac activities only modulate APG's (a) amplitude or (b) phase.

each heartbeat, which is defined as follows:

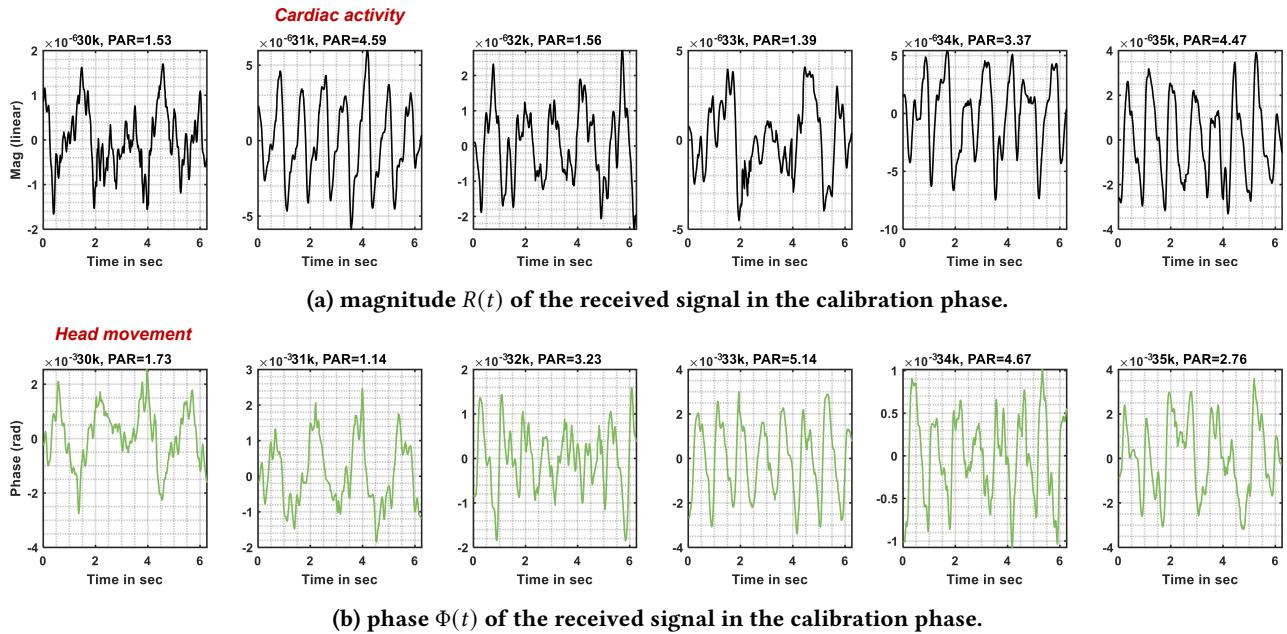
$$RMSSD = \sqrt{\frac{\sum_{i=1}^{N-1} (RR_i - RR_{i+1})^2}{N - 1}} \quad (7)$$

where RR_i is the peak time for the i th heartbeat.

Multi-tone APG. Coherent detection helps to retrieve the weak cardiac modulation signal from the raw ultra sound echo receptions. However, as shown in Figure 7, the mixed-down signal is noisy and thus cannot be directly used for HRV estimation. To address this challenge, we introduce a simple yet effective amendment to APG design, namely, *multi-tone APG*. Multi-tone APG is based on a key observation that the intensity of cardiac modulation is hypersensitive to the frequency of APG probing signals (§3.3). This allows us to manipulate the frequency of probing signals to improve the quality of cardiac signals. However, as the ear structure differs drastically with different users, the optimal frequency (*i.e.*, the frequency that leads to the highest signal quality) will change with users and wear conditions.

We design a *calibration* and *measurement* two-phase algorithm to obtain high-quality cardiac signals. In the calibration phase, we send multiple probing signals on N different frequencies simultaneously. The microphone will receive the reflection signals at N frequencies. For each of these N received signals, we extract both its phase profile and amplitude profile. By counting in the left ear channel and right ear channel, we get $4N$ signal profiles in total. Due to the frequency selectivity, some of these signal profiles can capture fine-grained cardiac activities well, while others cannot. Hence in the measurement phase, we pick two probing signals from N frequencies – one frequency that yields the best signal quality in the amplitude profile and another on the phase profile. We allocate all powers to these two frequencies and send our the probing signals for HRV measurement.

Characterizing the quality of received signals. We use the peak-to-average ratio (PAR) metric to characterize the quality of the received signal. Given a segment of the received signal, its PAR is defined as *the peak intensity within 0.58 Hz to 3.3 Hz frequency band over the average signal intensity*

(a) magnitude $R(t)$ of the received signal in the calibration phase.

Head movement

(b) phase $\Phi(t)$ of the received signal in the calibration phase.

Figure 8: The mixed down amplitude (upper row) and phase (bottom row) for a customized multi-tone APG signal that spans from 30 kHz to 35 kHz. The user is repeatedly moving her head during the 6 seconds of experiment. Due to the channel diversity, the cardiac activities are captured in some frequencies with high PAR (e.g. 31 kHz, mag) while the head movements are captured in some other frequencies with low PAR (e.g. 30 kHz, phase).

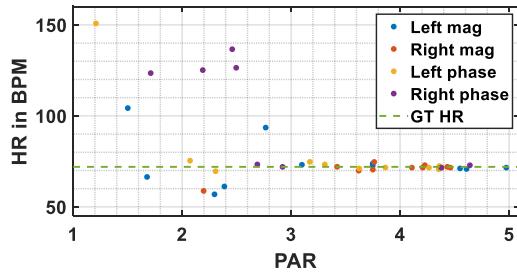
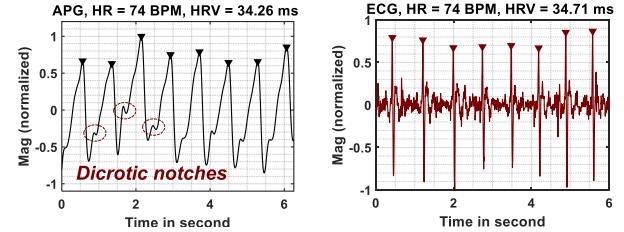


Figure 9: The final heart rate in BPM vs PAR. Each measurement comes from the amplitude or phase of various frequencies. The HR measurement results converge to ground truth (72 BPM) when the PAR is high.

within this band. A higher PAR indicates a stronger cardiac modulation signal, hence a better signal quality. We choose these particular frequency points because the human heart rate is typically within the range of 35 BPM to 200 BPM [57].

To validate the effectiveness of our signal quality metric, we transmit six probing signals concurrently with their frequencies spanning evenly from 30 KHz to 35 KHz. The user is asked to shake her head intentionally during the experiment to introduce interference. Figure 8 shows the received signal after coherent detection. As shown, the received signal with a higher PAR (e.g., PAR> 4) manifests a clear and stable heartbeat pattern, whereas both the phase and amplitude of those low PAR signals (e.g., PAR<2) are overwhelmed by head movements. We expand the number of transmissions



(a) final APG readings.

(b) ECG readings.

Figure 10: The final APG signal used in the measurement phase vs. chest ECG signal. The final APG signal reflects both heart rate and heart rate variability accurately. The dicrotic notches can even be observed in the waveform.

to 11 frequencies and get 44 signal profiles from the signals collected by both left- and right-headphone. We calculate heart rate based on each signal profile by autocorrelation [52], and plot their result in Figure 9. Evidently, the heart rate based on signal profiles with a higher PAR (i.e., PAR > 3.3) converge to the ground truth (72 BPM).

Physical intuitions of the APG frequency diversity. The observed APG frequency diversity is due to the fact that the ear canal is a frequency-selective fading channel, where modulated cardiac signals propagating through multiple paths can be coherently or non-coherently combined at different frequencies. Moreover, probing signals at different frequencies are diffracted and reflected differently by the ear canal, leading to another degree of frequency selectivity.

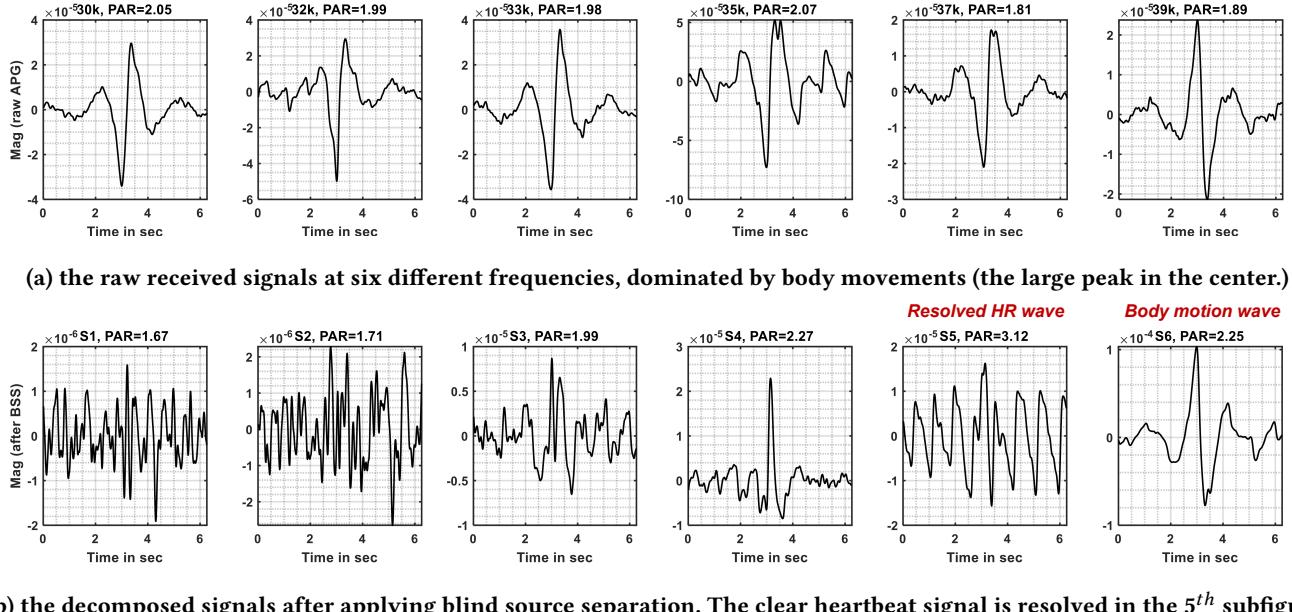
(b) the decomposed signals after applying blind source separation. The clear heartbeat signal is resolved in the 5th subfigure.

Figure 11: The mixed down amplitude (top row) and resolved source signals after the blind source separation pipeline (bottom row) for a customized multi-tone APG signal. There is a huge signal fluctuation caused by body motion in this experiment. The cardiac activities are floored by the body movement in all frequencies of the raw mixed-down APG. The blind source separation pipeline recovers and separates both cardiac activities and body movement. We can also see that the body movement modulation is around 10x larger than the cardiac activities modulation on the APG.

We measure the HRV using the received signal in the measurement phase (after calibration). Specifically, to calculate the RR interval time, *i.e.*, $RR_{i+1} - RR_i$, we use a peak finding estimator [12] to localize the peak of each heart pulse. Figure 10(a) shows an example of the received signal profile used for HRV measurement. Evidently, the cardiac activities are further pronounced and we can even see the blood pressure-related dicrotic notches in the raw APG signal. The HRV ground-truth shown in Figure 10(b) is obtained by a medical grade Zephyr BioHarness 3 [14] chest strap that was put on the subject's chest to collect ECG signals. The estimated HRV from APG approaches the ground truth HRV readings (the numbers on top of these two figures).

4.2 Heart Rate Monitoring in Motion

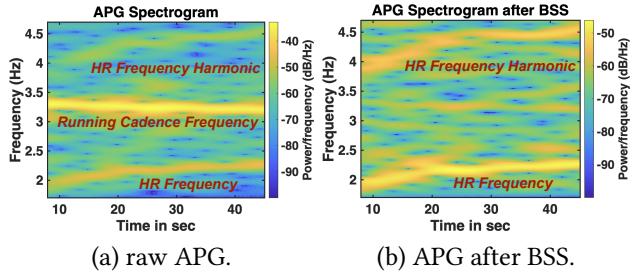
As described by the physical model (Figure 3), APG essentially measures the volumetric change in the ear canal. However, in practice, both head and body motions will squeeze the ear canals, altering their structures intermittently. Accordingly, these motion artifacts will be superimposed with the heartbeat signals, disturbing the cardiac signal detection. Since these body motion artifacts are orders of magnitude stronger than the cardiac activities, they will easily dominate the received signal, setting a strong barrier to the cardiac activity detection. The bottom left sub-figure in Figure 8 shows the head movement dominates the received signal. Although

the multi-tone APG design helps to find the frequencies that are hyper-sensitive to cardiac activities (*e.g.*, many subfigures in Figure 8), we are still seeing cases that the intensive body motions (*e.g.* walking) contaminate the echoes across all probing frequencies. To mitigate the impact of strong interference, we introduce a motion artifacts removal pipeline that takes advantage of our multi-tone APG design.

Blind source separation primer. Blind source separation (BSS) aims to recover each source signal from their mixtures, without the aid of information about the mixing process. BSS is equivalent to identifying the factorization of the N dimensional observations \mathcal{X} into a mixing channel \mathcal{A} and M dimensional source signals \mathcal{S} :

$$\mathcal{X} = \mathcal{A}\mathcal{S} \quad (8)$$

The matrices \mathcal{X} and \mathcal{S} represent the received signal mixture and the source signals. Note that in a BSS problem, we have no prior knowledge about the channel \mathcal{A} and the source signals \mathcal{S} . Instead, we only make a few general statistical assumptions on the source signals and the channel structure. \mathcal{A} explains various cross-statistics of the observations \mathcal{X} as an expansion of the corresponding diagonal cross-statistics of the sources \mathcal{S} . Let \mathcal{R}_x be the time averaged covariance matrix, $\mathcal{R}_x = \sum_t E[x(t)x(t)^H]$, where X^H denotes the Hermitian transpose of X . If \mathcal{R}_x is a determined matrix (*i.e.*,



(a) raw APG.

(b) APG after BSS.

Figure 12: The spectrogram of APG signals during a running session, (a) before and (b) after the blind source separation pipeline. The running dominates the raw APG signal whereas the heart rate and its harmonics are brought up after the blind source separation process.

$M \leq N$), we can write \mathcal{R}_x as follow:

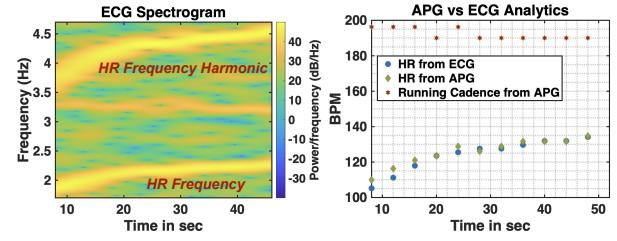
$$\mathcal{R}_x = \mathcal{A} \mathcal{R}_s \mathcal{A}^{-1}. \quad (9)$$

where \mathcal{R}_s is diagonal matrix. Note \mathcal{A} can be calculated by the eigenvalue decomposition of the covariance matrix \mathcal{R}_x . **Challenges in applying blind source separation.** The APG signal is modulated by multiple signal sources (e.g., heartbeat, head motion, body movement) in the ear canal. These sources are independent and decorrelated with each other. Therefore, the covariance matrix \mathcal{R}_x could be under-determined and we cannot extract the source signals from their mixture with blind source separation if we don't have enough independent observations (*i.e.*, $M > N$).

Blind source separation with multi-tone APG. We take advantage of the multi-frequency nature of APG probing signals to address this challenge. Specifically, the frequency bands of probing signals are statistically decorrelated due to the channel frequency diversity [49]. For instance, two separate frequency bands may capture totally different signal sources, as shown in Figure 8. Accordingly, the use of multi-frequency probing signals could significantly boost the number of independent observations and further make \mathcal{R}_x a determined covariance matrix. In our experiments, we found five frequencies are usually adequate to make \mathcal{R}_x an over-determined covariance matrix (*i.e.*, $M < N$).

For each earbud, we get an observation matrix X from the mix-down signals across all probing frequencies and apply an eigenvalue decomposition on \mathcal{R}_x to calculate the channel matrix \mathcal{A} . The source signals are recovered by $S = \mathcal{A}^{-1}X$.

Figure 11 gives an exemplary motion separation result. In this experiment, we send probing signals on six frequencies spanning from 30 kHz to 39 kHz. The user is asked to move her body *intensively* to introduce a large motion interference. Before applying the separation algorithm (Figure 11 (a)), we observe that the cardiac activities are overwhelmed by the body motions across all six frequency bands. In contrast, as we applying the blind source separation algorithm, we



(a) chest ECG ground-truth.

(b) analytics.

Figure 13: The spectrogram of chest ECG signal (ground-truth) collected by a Polar H10 and the analytics from the experiments. The ground-truth is collected at the same session in the experiment shown in Figure 12. APG not only tracks the heart rate accurately during the running session but also measures the running cadence.

can see the cardiac signal and body motion signal are disaggregated (Figure 11 (b)). Moreover, we observe that the body motion artifacts are about 10× stronger than the cardiac signals, which explains the result that the cardiac signal is not observable on the raw mix-down signals across all frequency bands. The resolved cardiac signal also has a significantly higher PAR than other sources.

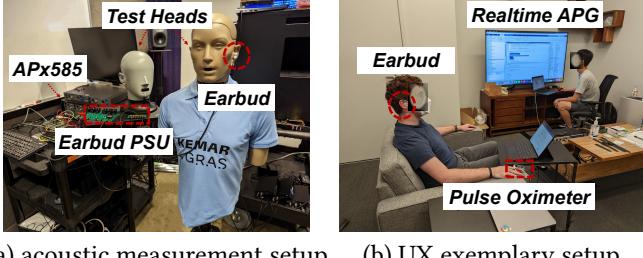
Figure 12 further shows the spectrogram of APG signals before and after applying blind source separation. The data is collected during a running session. Figure 13 (a) shows the ground truth heart rate measurement in the same running experiment from a Polar H10 ECG chest strap [6]. We can clearly see the running cadence (around 3.3 Hz) in Figure 12 (a) as well as two dimmer lines (around 2 Hz and 4 Hz) that indicate the user's heart rate frequency and its harmonics. The heart rate frequencies are significantly enhanced in SNR after the blind source separation, which align with the ground truth heart rate frequencies in Figure 13 (a). Figure 13 (b) shows the calculated heart rate and running cadence from APG and ECG. We can see that APG tracks the growth of heart rate during the running session accurately.

5 EVALUATION

In this section, we first describe the experiment setups (§5.1). We then conduct micro-benchmarks to examine the impact of different factors on the system performance (§5.2). Next, we run field studies to evaluate system performance (§5.3). Finally, we conduct UX studies to investigate the inclusiveness around this new technology (§5.4).

5.1 Experiment Setup

APG data collection. We collect APG data using four ANC earbud platforms, including a wired out CEA prototype and a CEB earbuds. The Product Team from our company has also built two dedicated ANC earbud prototypes that have a speaker and various number of microphones for our experiments. We use these prototypes to examine the impact



(a) acoustic measurement setup. (b) UX exemplary setup.

Figure 14: Experiment setups. We evaluated APG using IEC complied acoustic heads and conducted rigorous experiments designed by our UX research team.

of microphone positions on APG performance. The experiment setups are shown in Figure 14(a). We also build a power supply board (PSU) to supply the phantom power to these microphones. We use a RME Fireface UFX+ USB Audio Interface [20] as the analog-digital/digital-analog converter (AD/DA), where the microphone receiving gain is set to 40 dB. Finally, in order to measure the intensity of the transmitted acoustic signal at the DRP, we use IEC complied GRAS 45BB KEMAR Head & Torso [4], RA040, as well as Brüel & Kjær 5128 high-frequency head & torso simulator with an Audio Precision APx585 to measure the acoustics. The Brüel & Kjær 5128 dummy head is powered by a G.R.A.S 12BB power supply unit [17]. Additionally, we build an alternative test platform using a RME ADI-2 PRO FS R USB Audio Interface [19] and a TASCAM Series 8p Dyna microphone preamplifier [21] (with 40 dB gain) for cross validations.

Ground truth devices. We collect ground truth using multiple FDA-approved devices. The ground truth heart rate is collected by a Masimo MightySat Rx and a Nonin 9560 Onyx II Fingertip Pulse Oximeter [2]. The ECG and HRV ground truth are collected by Zephyr bioharness 3 and Polar H10 chest strap. All the ground truth devices provide Epoch & Unix timestamps. We synchronize the APG data and ground truth devices using the Epoch time stamps.

5.2 Micro-benchmark

5.2.1 APG ultrasound frequency limits. We first examine the impact of the frequency of probing signals on heart rate monitoring. In this experiment, we send a single-tone probing frequency at 10 different ultrasound band and measure the PAR of each received signal segment from the APG magnitude $R(t)$. The received ultrasound intensity at DRP is set to 45 dB SPL at 30 kHz. We repeat the experiment 10 times in each frequency setting. Recall that a higher PAR value indicates a better signal quality. A signal with a PAR > 3 manifests a clear and stable heartbeat pattern while PAR > 4 usually indicates good quality for the HRV measurement. The result is shown in Figure 15. We observe that the signal quality varies with the frequency of probing signals. Specifically, the PAR remains at a relatively low level (*i.e.*, around 3) when the

probing frequency is below 29 KHz. It then grows dramatically to over 4 as the probing frequency grows to 38 KHz. The PAR value then drops as the probing frequency grows further. Note from our theory analysis in Section 3.2, the PAR should increase as the frequency increases. However in real world settings, the PAR is affected by many other factors such as the frequency responses of speakers/microphones and/or the noise distribution over frequency.

5.2.2 APG ultrasound intensity requirements. Next we examine the PAR under different ultrasound intensity settings. In this experiment, we fix the frequency of the probing signal to 35 kHz. We then vary the intensity of the probing signal from 30 dB SPL to 75 dB SPL at DRP and measure the PAR in different intensity settings. The results are shown in Figure 16. We observe that the PAR grows gradually with increasing loudness (intensity) of the probing signal. Nevertheless, we can see the probing signal at 30 dB SPL intensity still achieves a decent PAR (around 4). Suggested by this result, we adopt 45 dB SPL as our default setting in the following experiments. It is worth noting that this intensity setting is far below the safety requirement: IRPA-specified ultrasound exposure limit at 30–40kHz frequency band is 110 dB SPL.

Implications on the ultrasound intensity. Note that 30 dB SPL is an extremely low loudness level. The normal loudness level for conversation is around 65 dB SPL [18]. The fact that APG works under extremely low ultrasound intensity brings several benefits for the system. Firstly, it ensures safe operation as APG can still work even 80 dB under the IRPA specified ultrasound exposure limit. Secondly, it introduces negligible power impact on the battery life as the speaker and microphone are always playing (for continuous noise cancellation) when the ANC/transparency mode is on.

5.2.3 Probing frequency separation. Recall that the success of blind source separation algorithm is determined by the de-correlation property of the multi-frequency APG. To determine a proper setup of the frequency spacing in our multi-frequency APG design, we send a pair of probing signals, with one of their frequency fixed to 35 kHz, and another changes from 35.1 kHz to 35.9 kHz. We then measure the correlation of the received signal at each frequency using Pearson Correlation Coefficient (PCC [28]). A lower PCC value indicates the higher likelihood that these two signals are de-correlated. The experiment is repeated five times in each setting. As shown in Figure 17, the PCC value declines with growing frequency gap between two probing signals. It reaches to the minimum value at 700 Hz (average PCC=0.23). On the other hand, given a fixed resource band (*e.g.*, 32 kHz to 38 kHz) and the total available power, the power and SNR of each APG tone declines with decreasing frequency spacing because the number of tones grows with decreasing frequency spacing. By jointly considering the de-correlation

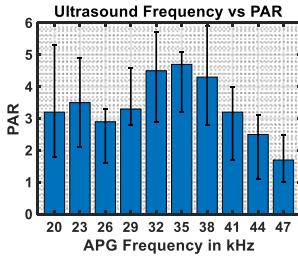


Figure 15: PAR under different APG frequencies.

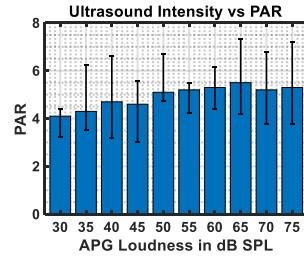


Figure 16: PAR under different signal intensities.

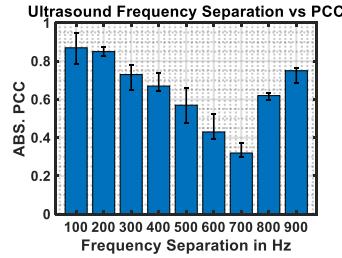


Figure 17: The correlations among frequency spacings.

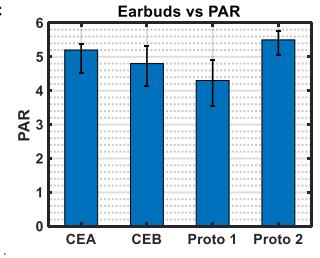
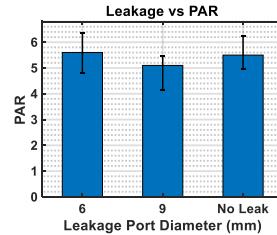


Figure 18: PAR vs various earbuds.



(a) leaky eartips.



(b) results.

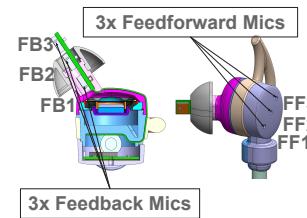
Figure 19: PAR vs emulated acoustic leakages.

coefficient and the power of each APG tone, we set the frequency spacing between adjacent APG signals to 1 KHz.

5.2.4 Impact of earphone heterogeneity. Next, we evaluate the signal quality captured by four different earphones. The experiment setups follow the previous benchmark studies. The results are shown in Figure 18. We observe that our prototype earphone 2 (Proto 2 in the figure) achieves the highest PAR (> 5.3), followed by CEA (around 5.1), CEB (around 4.8), and our prototype earphone 1(4.2). Nevertheless, even the worse performance earphones still achieve decent performances (recall that we can derive reliable heart rate from the received signals with a PAR > 3).

5.2.5 Impact of acoustic leakage. We then evaluate the impact of acoustic leakage on the received signal quality. In this experiment, we poke leakage ports on the eartips (Figure 19(a)) to emulate different seal conditions of the earphones. We measure the PAR of the received signal in different leakage port diameter settings (in mm). Figure 19(b) shows the result. We observe that the signal quality in the presence of different holes is on par with that in the absence of holes. Specifically, the average PAR in the absence of holes (*i.e.*, without signal leakage) is around 5.4. It then drops slightly to around 5.1 as we poke a 9mm hole on the eartips. The PAR value grows back to around 5.4 as the port diameter drops to 6 mm. Evidently, this result demonstrates that our design is not susceptible to moderate signal leakages.

5.2.6 Impact of microphone placement. We further examine the impact of microphone placement on the received signal quality (in PAR). To achieve this goal, our Product Team fabricated a new earphone with three feedback microphones



(a) Illustrations of 6 Mic locations.

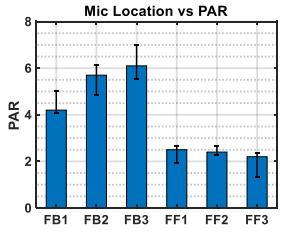


Figure 20: PAR vs various microphone locations.

and three feedforward microphones. Figure 20(a) shows the design sketch of this earphone prototype. We then send a multi-tone APG (the frequency grows from 32 kHz to 38 kHz with a 1 kHz spacing) and measure the PAR of the signal received by each of this six microphones. We repeat this experiment ten times and summarize the results in Figure 20(b). As shown, we find that in general the feedback microphones achieve better performance than feedforward microphones. This is expected since the feedback microphone are facing towards the inner ear canal and thus can capture stronger reflections. Among three locations, we find that FB3 that on top of the eartips achieves the best performance. However, it sticks out too much and hinders the comfort. Hence we adopt FB2 as our default settings for field studies.

5.3 Field Studies

5.3.1 Heart rate monitoring accuracy. We first examined the heart rate (HR) monitoring accuracy in four different states: stationary, active, stationary with music playing, and running. In the stationary mode, we invited 153 human subjects and measured each participant's HR for one minute using our proposed solution. The participant was also required to wear a FDA-approved Nonin 9560 Onyx II Fingertip Pulse Oximeter to collect ground-truth HR. In the active mode, we repeated the above experiments on 37 users while they are walking or watching news with natural body rotation and movement. In the stationary with music playing mode, we played 20 different music tracks with diverse frequency properties at 70 to 75 dB SPL to evaluate the impact of music playing on APG. In the studies where participants ran,

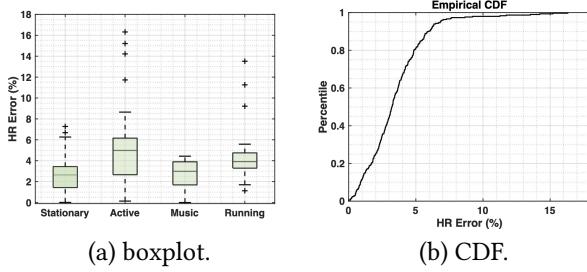


Figure 21: Statistics for HR monitoring across activities among 153 human subjects.

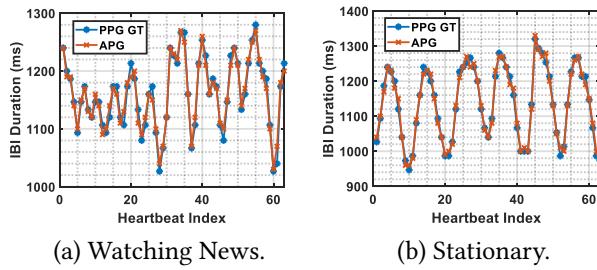


Figure 22: Example IBI signals for two activities.

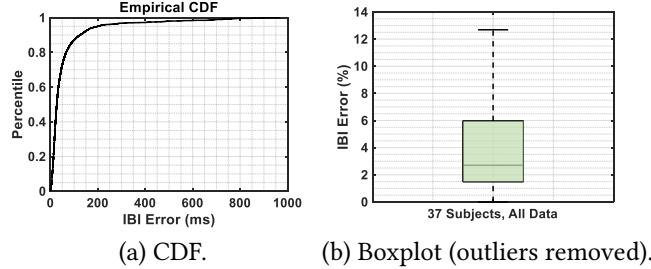


Figure 23: Statistics for IBI from 8771 RR intervals out of 37 participants.

52 one-minute running sessions with two users were conducted over the course of a few weeks. The participants were required to wear a Polar H10 ECG chest strap for HR ground-truth. The results are summarized in Figure 21(a).

Overall, we observe that APG achieves consistently low HR error rate in two stationary modes. The median HR error in stationary and stationary with music playing is 2.64% and 2.99%, respectively. The median HR error grows to 3.97% and 5.01% in running and active mode because human motion can occasionally add noise to the heartbeat signal detection. We further plot the CDF of the overall 153 subjects HR error across all scenarios in Figure 21(b). The 85% percentile error rate is 5.40%. The median error rate is 3.21%.

5.3.2 HRV accuracy. Next, we evaluated the accuracy of HRV estimation based on APG. We invited 37 volunteers to measure their HRV in a living room. The volunteers were asked to either watch news or take a rest on sofa. Over 8700 RR intervals have been measured from these volunteers. We then compute the interbeat interval (IBI) between each

consecutive RR intervals and compared it to those values measured by an FDA-approved Nonin 9560 Onyx II Fingertip Pulse Oximeter. Figure 23(a) shows the CDF of the IBI error (compared to the ground-truth collected by the dedicated device mentioned above). We observe that the median and 85% quantile IBI error is 26.70 ms (equivalent to 2.70% error shown in Figure 23(b)) and 86.74 ms, respectively. This result meets the requirement set by our health team (IBI median error should be < 5%). In Figure 22 we show two snapshots of the estimated IBI duration when the volunteer is watching news (with occasional body rotation/movement) or stays stationary. We also plot the PPG results in these two snapshots for comparison. Evidently, the IBI duration reported by APG is very close to that collected by the dedicated PPG sensor in all cases. These results clearly demonstrate the efficacy and efficiency of APG in HRV measurement.

5.4 Validating the Technology

We collaborated with the UX research team in our company and designed rigorous UX experiments to validate APG.

5.4.1 UX Study 1. Study 1 consists of 34 participants (20 male-identifying, 14 female-identifying) spread across age ranges (21-59), with 50 percent of the study population falling between ages 21-29. Data was gathered with participants seated and 1) breathing normally or 2) following specific box-breathing instructions.

Study 1 procedure: Participants answered survey questions to understand their attitudes and thoughts around health, technology, health-tech, and health-tech pain-points. They were then asked to describe their current heart rate measurement flow (the hows, whens, and whys of measuring their heart rate). After that, the participant donned headphones enabled with the APG technology but were not told anything about how the headphones were to be used. Then heart rate data was collected with different eartip sizes (S/M/L). Participants filled out questions about the experience of using the headphones. They then figured out that heart rate was measured by the earbuds upon the proctors intentionally displaying data on a TV in the room. After the participants gave their feedback, one of the proctors gave a brief overview of the technology so experimenters could understand any questions/concerns participants had about the technology. The quantitative results are shown in Figure 24 (a). Overall, our primary takeaway is that APG is insensitive to different sizes of ears and eartips.

5.4.2 UX Study 2. Study 2 consists of 37 participants (25 male-identifying, 12 female-identifying) spread across age ranges (21-69). The age groups in this study are more diverse than Study 1, a key goal to ensure diverse representation in our population. Overall, we have 5 people in the 50+ category, 16 people in the 30-39 category, 7 people in the 40-49 category and 9 people in the 21-29 category. Similarly to

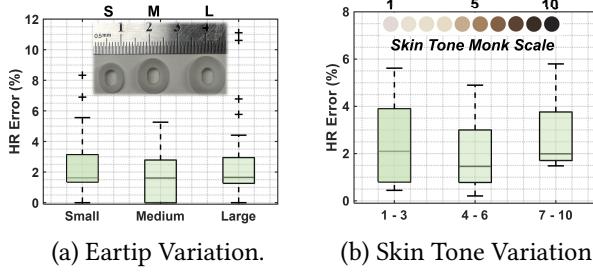


Figure 24: Quantitative results from UX studies.

UX Study 1, participants self-report eartip size, but in this study, the eartips were medium by default. Where Study 2 primarily differed from Study 1 are in the specific request for participants to self-select their skin tone using the Monk Scale [1, 16] and with participants complete more activities. Given classical PPG's sensitivity to skin tone, our primary goal is to understand the diversity amongst this specific study population. The breakdown of what participants chose is as follows: 13/37 chose 1 - 3, 19/37 chose 4 - 6, and 5/37 chose 7 - 10. Although we don't have full coverage using the Monk scale, we feel confident our results will translate to a larger more diverse population. The results are shown in Figure 24 (b). APG achieves consistent low heart rate error across all three groups, demonstrating its inclusiveness to skin tone.

6 RELATED WORK

Earable sensing with dedicated sensors. There is a growing trend of putting embedded sensors in the earbuds for physiological sensing. Heart rate, respiratory rate, and blood pressure can be monitored using ECG, PPG, etc. [3, 8, 10, 11, 23, 34, 37]. For instance, Bui *et al.* developed an in-ear system to measure blood pressure using PPG sensors [24]. Anh *et al.* proposed to customize an in-ear sensor to measure brain activities [41]. Roddiger *et al.* developed a respiration rate monitoring system using the embedded IMU [48]. There is also an open-source multi-sensor integrated research platform, eSense, for earable computing research [7]. These sensing systems and platforms rely on dedicated sensors that add weight, compromise the acoustic design, and incur a higher cost and power consumption. In contrast, APG leverages commodity ANC headphones to monitor heart rate, without relying on any new sensor hardwares.

Sensing with headphone transducers. There are also various works leveraging the microphones and speakers on headphones for human-centric sensing [25, 30, 36, 39, 58, 59]. For instance, EarGate [32] demonstrates the feasibility of user identification by tracking the unique user gate through earphones. Likewise, OESense [39] takes advantage of in-ear microphone to sense a set of human activities including step counting, activity recognition, and hand-to-face gesture interaction. hEARt [25] further leverages the in-ear

microphone to monitor heart rate in an unconstrained environment. HeadFi [30] retrofits the built-in speakers in headphones to realize various of smart applications. EarHealth [36] detects middle ear diseases by sending a 20 Hz to 6 kHz (audible) acoustic probing tone. A recent work Earmonitor [55] sends a 16 kHz - 21 kHz chirp for heart rate monitoring. However, the audible chirp signals will distort the music (music spectrum spans from 20 Hz - 20 kHz) as well as complicate the ANC operation. Chirp signals also require advanced digital signal processing techniques to extract useful information from the received signal. This complexity may increase the computational requirements, power consumption, and implementation costs. This is especially true in the TWS earbuds silicon design. In contrast to these system designs, APG spearheads a simple yet unique active probing signal using true ultrasound, yields a PPG like cardiac waveform for both heart rate and heart rate variability monitoring. It further takes advantage of frequency diversity to enhance the heart rate monitoring accuracy in unconstrained environment.

7 CLOSING THOUGHTS

We have presented the design, implementation, and evaluation of APG, a novel cardiac monitoring modality for ANC headphones. With a software upgrade, APG adds the cardiac monitoring feature to all ANC headphones with negligible battery life impact. APG is resilient to music playback, body motions, skin tone, and different sizes of eartips. Our large-scale field studies show that APG achieves a median error of 3.21% and 2.70% on HR and HRV monitoring, respectively.

Our future research will focus on developing more rigorous motion artifacts removal pipeline to ensure the high performance of APG in various mobile user scenarios, such as hiking, weightlifting, squat, boxing, HIIT (high-intensity interval training). On the other hand, We envision that APG could enable many other exciting health research as the APG waveform contains lots of nuances beyond just HR and HRV. For example, one may leverage the dicrotic notches shown in the APG waveform (Figure 1) to measure the blood pressure through pulse wave analysis [22].

8 ACKNOWLEDGMENTS

We sincerely thank our MobiCom shepherd and reviewers for their constructive comments and feedback, and Shwetak Patel, for his invaluable critiques and reviews. Longfei Shangguan is supported by a Google Research Scholar Award.

REFERENCES

- [1] Improving skin tone representation across google. Website, 2012.
- [2] Onyx ii model 9560 finger pulse oximeter. Website, 2012.
- [3] BOSE QC-35 Wireless Headphones. Website, 2020.
- [4] Gras 45bb kemar head & torso. Webpage, 2020.
- [5] Gras ra0404 prepolarized hi-res ear simulator. Webpage, 2020.
- [6] Polar h10 - the gold standard. Webpage, 2020.
- [7] A research space for earable computing. Webpage, 2020.
- [8] Samgsung Galaxy Buds. Website, 2020.
- [9] Wireless ecg, spo2, ptt and heart rate monitor reference design for medical and consumer wearables. Website, 2020.
- [10] Apple airpods. Website, 2021.
- [11] Earable: Helping the world sleep better. Webpage, 2022.
- [12] Find local maxima. Website, 2022.
- [13] Sawtooth wave. Website, 2022.
- [14] Zephyr™ performance systems. Website, 2022.
- [15] Apx58x b series audio analyzers. Website, 2023.
- [16] Developing the monk skin tone scale. Website, 2023.
- [17] Gras 12ba, 12be and 12bb ccp power modules with teds support. Website, 2023.
- [18] How many decibels does a human speak normally. Website, 2023.
- [19] Rme adi-2 pro fs r black edition. Website, 2023.
- [20] Rme fireface ufx+. Website, 2023.
- [21] Tascam series 8p dyna. Website, 2023.
- [22] Alberto P Avolio, Mark Butlin, and Andrew Walsh. Arterial blood pressure measurement and pulse wave analysis—their role in enhancing cardiovascular assessment. *Physiological measurement*, 2009.
- [23] Ananta Narayanan Balaji, Andrea Ferlini, Fahim Kawsar, and Alessandro Montanari. Stereo-bp: Non-invasive blood pressure sensing with earables. In *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications*, 2023.
- [24] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, et al. ebp: A wearable system for frequent and comfortable blood pressure monitoring from user's ear. In *The 25th annual international conference on mobile computing and networking*, 2019.
- [25] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. heart: Motion-resilient heart rate monitoring with in-ear microphones. *arXiv preprint arXiv:2108.09393*, 2021.
- [26] Ishan Chatterjee, Maruchi Kim, Vivek Jayaram, Shyamnath Gollakota, Ira Kemelmacher, Shwetak Patel, and Steven M Seitz. Clearbuds: wireless binaural earbuds for learning-based speech enhancement. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 2022.
- [27] Romit Roy Choudhury. Earable computing: A new area to think about. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, 2021.
- [28] Israel Cohen, Yiteng Huang, Jingdong Chen, Jacob Benesty, Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. *Noise reduction in speech processing*, 2009.
- [29] Francis A Duck. Medical and non-medical protection standards for ultrasound and infrasound. *Progress in biophysics and molecular biology*, 2007.
- [30] Xiaoran Fan, Longfei Shangguan, Siddharth Rupavatharam, Yanyong Zhang, Jie Xiong, Yunfei Ma, and Richard Howard. Headfi: bringing intelligence to all headphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021.
- [31] Xiaoran Fan and Trausti Thormundsson. Design earable sensing systems: Perspectives and lessons learned from industry. In *EarComp*, 2023.
- [32] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. Eargate: gait-based user identification with in-ear microphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021.
- [33] Hannes Gamper, Sakari Tervo, and Tapio Lokki. Head orientation tracking using binaural headset microphones. In *Audio Engineering Society Convention 131*. Audio Engineering Society, 2011.
- [34] Francis Roosevelt Gilliam III, Robert Ciesielski, Karlen Shahinyan, Pratistha Shakya, John Cunsolo, Jal Mahendra Panchal, Bartłomiej Król-Józaga, Monika Król, Olivia Kierul, Charles Bridges, et al. In-ear infrasonic hemodynography with a digital health device for cardiovascular monitoring using the human audiome. *npj Digital Medicine*, 2022.
- [35] Jon D Hendrix, Jeffrey D Alderson, Chin Huang Yong, and Ryan A Hellman. Feedback howl management in adaptive noise cancellation system, 2019. US Patent 10,290,296.
- [36] Yincheng Jin, Yang Gao, Xiaotao Guo, Jun Wen, Zhengxiong Li, and Zhanpeng Jin. Earhealth: an earphone-based acoustic otoscope for detection of multiple ear diseases in daily life. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 2022.
- [37] Fahim Kawsar, Chulhong Min, Akhil Mathur, Alessandro Montanari, Utku Günay Acer, and Marc Van den Broeck. esense: Open earable platform for human sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, 2018.
- [38] Wolfgang Klippel. Nonlinear large-signal behavior of electrodynamic loudspeakers at low frequencies. *Journal of the Audio Engineering Society*, 1992.
- [39] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. Oesense: employing occlusion effect for in-ear human sensing. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021.
- [40] Monique Moderiano, Maureen McEvoy, Jessie Childs, and Adrian Esterman. Safety of ultrasound exposure: knowledge, attitudes, and practices of australasian sonographers. *Journal of Diagnostic Medical Sonography*, 2018.
- [41] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei-Kashani, Ann C Halbower, and Tam Vu. A lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, 2016.
- [42] Shahriar Nirjon, Robert F Dickerson, Qiang Li, Philip Asare, John A Stankovic, Dezhi Hong, Ben Zhang, Xiaofan Jiang, Guobin Shen, and Feng Zhao. Musicalheart: A hearty way of listening to music. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, 2012.
- [43] Jun Nishimura and Tadahiro Kuroda. Eating habits monitoring using wireless wearable in-ear microphone. In *2008 3rd International Symposium on Wireless Pervasive Computing*. IEEE, 2008.
- [44] International Non-Ionizing Radiation Committee of the International Radiation Protection Association et al. Interim guidelines on limits of human exposure to airborne ultrasound. *Health Physics*, 1984.
- [45] Sebastian Passler and Wolf-Joachim Fischer. Food intake activity detection using a wearable microphone system. In *2011 Seventh International Conference on Intelligent Environments*. IEEE, 2011.
- [46] Nhat Pham, Tuan Dinh, Zohreh Raghebi, Taeho Kim, Nam Bui, Phuc Nguyen, Hoang Truong, Farnoush Banaei-Kashani, Ann Halbower, Thang Dinh, et al. Wake: a behind-the-ear wearable system for microsleep detection. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, 2020.

- [47] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. Sensing with earables: A systematic literature review and taxonomy of phenomena. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2022.
- [48] Tobias Röddiger, Daniel Wolfgram, David Laubenstein, Matthias Budde, and Michael Beigl. Towards respiration rate monitoring using an in-ear headphone inertial measurement unit. In *EarComp*, 2019.
- [49] Daniel Rouseff, John A Flynn, James A Ritcey, and Warren LJ Fox. Acoustic communication using time-reversal signal processing: Spatial and frequency diversity. In *AIP Conference Proceedings*. American Institute of Physics, 2004.
- [50] Fred Shaffer and Jay P Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, 2017.
- [51] SV Shilko, SP Salivonchik, et al. Calculation of pulse wave parameters with account of blood vessel deformation. *Russian Journal of Biomechanics*, 2001.
- [52] Robert R Sokal and Neal L Oden. Spatial autocorrelation in biology: 1. methodology. *Biological journal of the Linnean Society*, 1978.
- [53] Phyllis K Stein, Matthew S Bosner, Robert E Kleiger, and Brooke M Conger. Heart rate variability: a measure of cardiac autonomic tone. *American heart journal*, 1994.
- [54] Wei Sun, Franklin Mingzhe Li, Benjamin Steeper, Songlin Xu, Feng Tian, and Cheng Zhang. TeethTap: Recognizing discrete teeth gestures using motion and acoustic sensing on an earpiece. In *26th International Conference on Intelligent User Interfaces*, 2021.
- [55] Xue Sun, Jie Xiong, Chao Feng, Wenwen Deng, Xudong Wei, Dingyi Fang, and Xiaojiang Chen. Earmonitor: In-ear motion-resilient acoustic sensing using commodity earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2023.
- [56] Conny MA van Ravenswaaij-Arts, Louis AA Kollee, Jeroen CW Hopman, Gerard BA Stoelinga, and Herman P van Geijn. Heart rate variability. *Annals of internal medicine*, 1993.
- [57] Joseph C Volpe Jr. Heart rate monitor for controlling entertainment devices, April 8 2008. US Patent 7,354,380.
- [58] Zi Wang, Yili Ren, Yingying Chen, and Jie Yang. Toothsonic: Earable authentication via acoustic toothprint. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2022.
- [59] Metin C Yarici, Harry J Davies, Takashi Nakamura, Ian Williams, and Danilo P Mandic. Hearables: In-ear multimodal brain computer interfacing. In *Brain-Computer Interface Research*. Springer, 2021.