# Exploring the Feasibility of Remote Cardiac Auscultation Using Earphones

Paper #263: Revise-and-Resubmit

## Abstract

The elderly over 65 accounts for 80% of COVID deaths in the United States. In response to the pandemic, the federal, state governments, and commercial insurers are promoting video visits, through which the elderly can access specialists at home over the Internet, without the risk of COVID exposure. However, the current video visit practice barely relies on video observation and talking. The specialist could not assess the patient's health conditions by performing auscultations.

This paper tries to address this key missing component in video visits by proposing Asclepius, a hardware-software solution that turns the patient's earphones into a stethoscope, allowing the specialist to hear the patient's fine-grained heart sound (*i.e.*, PCG signals) in video visits. To achieve this goal, we contribute a low-cost plug-in peripheral that repurposes the earphone's speaker into a microphone and uses it to capture the patient's minute PCG signals from her ear canal. As the PCG signals suffer from strong attenuation and multi-path effects when propagating from the heart to ear canals, we then propose efficient signal processing algorithms coupled with a data-driven approach to de-reverberate and further correct the amplitude and frequency distortion in raw PCG receptions. We implement Asclepius on a 2-layer PCB board and follow the IRB protocol to evaluate its performance with 30 volunteers. Our extensive experiments show that Asclepius can effectively recover Phonocardiogram (PCG) signals with different types of earphones. The objective blind testing and subjective interview with five cardiologists further confirm the clinical efficacy and efficiency of our system. PCG signal samples, benchmark results, and cardiologist interviews can be found at an anonymous link https://asclepius-system.github.io/

## 1 INTRODUCTION

Imagine an old man approaching his eighty, suffering from chronic diseases and living tens of miles away from the nearest medical center. *Video visit* that allows him to access specialists timely from his own home could mean life or death for him [17]. Due to the coronavirus, going to a clinic, a hospital, or even taking a standard check-up may put the elderly in danger. We thus have witnessed a rapid growth of video visit services in the past few years. Even in the age of post-pandemic, health organizations are promoting video visit to avoid unnecessary emergency department visits and prolonged hospitalizations [26, 66].

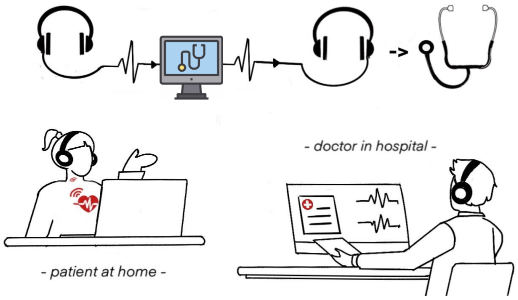


**Figure 1: Asclepius empowers the specialist to hear the patient's heart sound during a video visit.**

While video visit has opened the door for the elderly to maintain access to specialists at home, the current practice of video visit is far less effective compared to physical visit because evaluating the patient's health condition remotely is challenging: specialists observe via video and communicate symptoms by talking to the patient, they however cannot perform *auscultation* – an indispensable physical examination (PE) methodology – to make therapeutic decisions.

Cardiac auscultation is the most crucial physical examination among others [56]. It is performed to examine the circulatory system by listening to the heart sound (*i.e.*, PCG signal) emanating from the human heart. Although major pharma providers have rolled out plenty of in-home digital stethoscope that allows patients to measure their PCG signals at home and synchronize their data with specialists through Wi-Fi or Bluetooth connection, these devices are usually pricey (*e.g.*, Thinklabs One digital stethoscope [74] costs $499 USD) and difficult to operate for the elderly. More importantly, even with access to these devices, patients lack professional training would not be able to place a stethoscope at the right place for heart sound collection.

This paper explores the feasibility of designing a remote auscultation solution for video visits. The proposed solution should satisfy the following requirements. **High accuracy**. The solution should be able to detect both coarse-grained heart rate variation (HRV) and fine-grained cardiac features (*e.g.*, S1, S2 sound, and possible heart murmurs) that are essential to cardiac auscultation. **Easy to operate**. The proposed system should also be easy to operate, allowing specialists to take remote cardiac auscultation with minimum patient intervention. **Low-cost**. The proposed system should also be low-cost (*e.g.*, less than $10 USD) so that it can scale to serve large populations rapidly and unobtrusively.

We achieve the above goals by proposing Asclepius, a hardware-software solution that turns the speaker transducer on the patient's earphone into a stethoscope and uses it to continuously monitor the acoustic cardiopulmonary signals from the patient's ear canal, with no explicit patient intervention. Our solution works with everyday earphones (*e.g.*, those earphones cost a few US dollars) and requires neither dedicated in-ear microphones nor IMU sensors (*e.g.*, accelerometer) that are only available on those pricey ANC earphones.

Developing Asclepius faces multiple challenges.

• First, unlike the dedicated stethoscope where the diaphragm is placed right above the heart with gentle force to best capture the heart sound (*i.e.*, PCG signals) [42], the PCG signals captured by an earphone experience significant attenuation and frequency distortion when propagating through the human bones, muscles, fat, and skins before arriving at the human ear [43]. Accordingly, these PCG receptions tend to be very weak and thus are likely to be buried by ambient noises and human organ artifacts.

• Second, although using speaker as a microphone is feasible due to their structure reciprocity [20, 24, 59], capturing PCG signals with an earphone's speaker is still challenging because the earphone speaker is optimized for signal emission, not for signal absorption. Accordingly, when the weak PCG signal arrives at the speaker's diaphragm, only a small portion of this signal will be transformed into a voltage signal. This weak voltage signal is unlikely to maintain the fine-grained PCG features such as S1 and S2 heart sound components.

• Third, an acoustic signal will get diffracted, reflected, and absorbed when propagating from the audio cables to the pairing device. The proportion of signal being absorbed by the pairing device is affected by the *mismatch* between the two impedances. The conventional offline impedance matching can not be applied to our problem because both the earphone's impedance and the pairing device's impedance are unknown. They also change dramatically with hardware type, form factor, and material. To cope with these dynamics, it is essential to conduct an online, automatic impedance matching.

To address the above challenges, Asclepius contributes a novel hardware plugin module coupled with an efficient software signal processing pipeline that works hand in hand to capture, amplify, and further correct the distortion of raw PCG receptions, as shown in Figure 2.

• Our hardware plugin turns the earphone's speaker into an agile microphone and uses this *microphone* to capture the minute PCG signal at the ear canal. It then amplifies this PCG signal and denoises the strong noises in the analog domain with a low-power analog circuit. To ensure the PCG signals can be delivered to the pairing device with minimum signal reflections, we further design a programmable impedance circuit and propose a feedback-loop-based control algorithm to balance the impedance between the earphone and the pairing device automatically, without any human intervention.

• Upon receiving the PCG signals, our signal processing pipeline running on the pairing device de-reverberates the raw PCG reception, segments them into heart cycles, and then corrects the frequency and phase distortion caused by the multi-path effect when the PCG signal propagates inside the human body. The output is sent to the specialist hereafter.

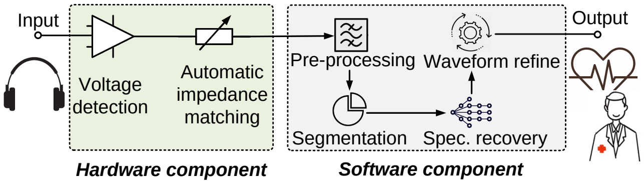


Figure 2: Overview of Asclepius.

We implement Asclepius's hardware on a 2-layer printed circuit board (PCB). The total hardware cost is $5 USD. The accompanying software processes, encompassing preprocessing, segmentation, spectrogram recovery, and waveform refinement, are implemented on a local pairing device (a laptop) and managed on a cloud server. We evaluate Asclepius using 12 pairs of commodity earphones. The results based on 30 volunteers of different ages, genders, and BMIs show that Asclepius achieves consistently high PCG signal recovery accuracy, with 1.17% average Root Mean Squared Error (RMSE) compared to ground-truth PCG signals. We also play 20 types of pathological PCG signals using a speaker attached to one end of a pork belly. These signals propagate through the 40cm pork belly and arrive at the earphones placed on the other end, experiencing strong multi-path fading. The emulation results show that Asclepius can effectively recover the multi-path distorted pathological PCG signals.

To examine the clinical efficacy of Asclepius, we invite five cardiologists to participate in a two-phase UX study. The blind testing in the first phase shows the diagnosis accuracy based on Asclepius's data is consistent with that based on the stethoscope's output across all five cardiologists. The subjective evaluation in the second phase shows that all five cardiologists can identify the S1 and S2 heart sounds and the pathological heart murmurs from Asclepius's recordings. They also believe Asclepius could serve as a valuable tool for remote visits, providing a trusting relationship between patients and clinicians.

**Claims**. Different from a prior poster [2], this research project demonstrates the feasibility of using commodity earphones to detect fine-grained PCG signals from the ear canal. The preliminary results are promising and the feedback from cardiologists is also positive. On the other hand, we acknowledge that Asclepius can only be used as a prescreening device to assist video visits; the current prototype cannot replace the dedicated stethoscope for a physical examination before undergoing a rigid, comprehensive clinic study. The reasons are twofold. First, the current testing cases are still very limited, and we may still face domain gaps between different subjects, which could affect the signal reconstruction performance. Second, the emulation of in-body transmission based on pork belly may not reflect the signal propagation inside human bodies fairly. To close the gap, we have been consulting clinicians during the development of Asclepius and are currently working closely with ABC (anonymized for double-blind review) medical center to initiate clinic studies.

**Contributions and roadmap**. Overall Asclepius makes the very first step toward remote auscultation, opening the door to efficient video visits. Moreover, we believe this project will spark novel ideas on heart sound sensing, pushing the whole field moving forward. The rest of the paper is organized as follows. We present the background and motivation (§2), followed by the hardware (§3) and software design (§4). We then describe the system evaluation in §5. Section 6 describes the related work. Section 7 concludes.

# 2 BACKGROUND AND MOTIVATION

In this section, we first explain cardiac auscultation and its significance in clinic pre-screening. We then discuss the challenges and opportunities for remote auscultation.

## 2.1 Cardiac Auscultation Primer

Cardiac auscultation [1] was recognized as a cornerstone for physical examination and medication since the early 19th century. Medical professionals such as well-trained physicians or specialists could assess a patient's cardiovascular activities and make objective therapeutic decisions by placing a stethoscope [73] on the chest of the subject and examining the internal sounds. A stethoscope is a sound system that can capture fine-grained Phonocardiogram (PCG) signals, including the first heart sound (S1), the second heart sound (S2) as well as higher pitch sounds such as heart murmurs generated from the closure and open of the heart valves and vessels when blood goes through heart atrium and ventricle.

Cardiac auscultation based on a stethoscope is low-cost, easy to operate, and user-friendly. As such, it has been adopted worldwide and serves as a standard for the nursing practice [23]. *Although many advanced technologies such as the electrocardiogram (ECG) and echocardiography have been invented for fine-grained cardiovascular activity monitoring, cardiac auscultation with a stethoscope is still an irreplaceable option in nursing practice. It not only helps to find a path towards diagnosis but also serves as an opening to a trusting, caring relationship between patients and specialists [18, 55].*

## 2.2 Remote Auscultation: Opportunities

The demand for video visits remains strong after the pandemic [58]. However, cardiac auscultation is still a daunting task in video visits. Recently, the proliferation of mobile devices may break this stalemate. For instance, prior works [34, 46] have demonstrated the potential of using smartphones to capture heart sounds. However, like the predicament faced by digital stethoscopes, without the necessary nursing practice, it is challenging for a patient to put the smartphone on the correct chest locations for heart sound capturing. Besides, smartphones adopt omnidirectional microphones to capture human speech, which makes them susceptible to motion artifacts and ambient noises during auscultation.

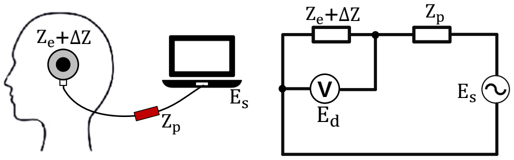


**Figure 3: Impedance variation measurement. The change of impedance will alter the voltage $E_d$.**

**Earphones as a stethoscope**. Compared to smartphones, earphones hold many unique advantages in cardiac auscultation.

- **Suffer from less ambient noises**. When putting on the earphone, the ear cup, ear canal, and eardrum will couple together, forming a hermetic space [50]. The ear cup will block ambient noises from entering the ear canal. Meanwhile, the heart sound will be amplified in the ear canal due to the occlusion effect [45].
- **Low system cost**. Every single pair of earphones has two speaker transducers for music playback. Due to structure reciprocity [71], these speaker transducers can be used as microphones to capture acoustic signals inside the ear canal. This leaves us with a cost-effective solution for auscultation.
- **Easy to operate**. The earphone-based solution allows the specialist to take cardiac auscultation online without any patient intervention.

# 3 ASCLEPIUS'S HARDWARE DESIGN

The PCG signals propagate through the human body to arrive at the ear canal. The earphone speaker's diaphragm responds to these signals, and a weak voltage signal is generated and then offloaded to the pairing device (*e.g.*, a desktop or a tablet that the patient uses to talk to the specialist) through the audio chain. Asclepius explores this opportunity to enable remote cardiac auscultation. In this section, we first model the relationship between the impedance variation and the inductive voltage signal. We then propose a low-power circuit to detect this voltage signal.

## 3.1 A Theoretical Model

When an earphone connects to a pairing device, a constant, bias voltage signal $E_s$[1] will go through the earphone's audio jack, arriving at the earphone's diaphragm. As shown in Figure 3, let $Z_e$ and $Z_p$ be the impedance of the earphone and the signal detection circuit (which will be introduced in next section), respectively; $\Delta Z$ is the earphone's impedance variation due to the PCG signal. $Z_p$, $Z_e$ are serially connected with each other, forming a voltage division circuit. Based on Ohm's law, we have:

$$E_d = \frac{Z_e + \Delta Z}{Z_e + Z_p + \Delta Z} \cdot E_s \tag{1}$$

[1] It is reasonable to request both the patient and the specialist to keep silent during auscultation. Hence $E_s$ would not change over the course of PCG signal detection.

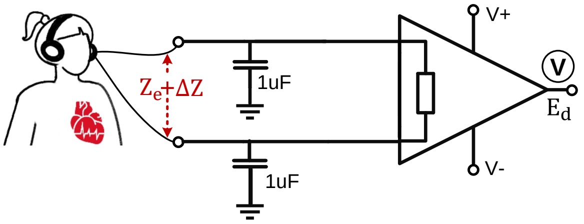


**Figure 4: Schematic of the voltage detection circuit.**

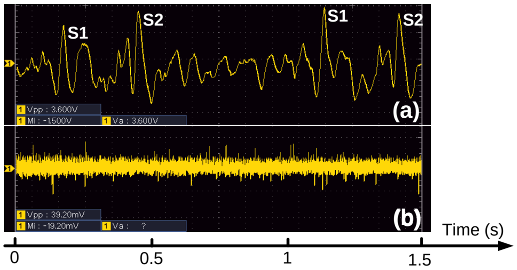


**Figure 5: Inductive voltage signal (a) with and (b) without the proposed detection circuit.**

Since the impedance variation $\Delta Z$ caused by heartbeats is orders of magnitude smaller than $Z_e + Z_p$, the above equation can be simplified as:

$$E_d = \frac{Z_e + \Delta Z}{Z_e + Z_p} \cdot E_s \tag{2}$$

Since both $Z_e$ and $Z_p$ are constant values, the voltage signal $E_d$ varies in proportion to $Z_e + \Delta Z$. Accordingly, it is feasible to detect PCG signals by tracking the voltage signal $E_d$. However, since the PCG signal is very weak after propagating along the human body, the variation of voltage signal $E_d$ due to PCG signals would be very subtle.

## 3.2 Inductive Voltage Detection Circuit

We propose a low-power detection circuit to detect $E_d$ from the patient's left ear transducer since the human heart is relatively closer to the left ear [15]. The right-ear channel is reserved for sound playback.

Figure 4 shows the schematic of this circuit. It consists of a low-noise operational amplifier and peripheral circuits (*i.e.*, a set of passive resistors and capacitors). The amplifier connects to the left-ear speaker transducer through a 3.5mm audio jack. We pick the amplifier with good frequency response on low frequencies (*e.g.*, < 1kHz) to avoid extra frequency distortion on PCG signals. We then add two identical bypass capacitors (1uF) before the amplifier to filter out high-frequency noises above the frequency of PCG signals. The equivalent series resistances [22] of these capacitors also improve the common-mode rejection ratio of the amplifier, ensuring a high amplification gain. Recall that the inductive voltage signal $E_d$ varies in proportion to $Z_e + \Delta Z$ (Equation 2), not $\Delta Z$ alone. Hence we are expected to see a strong common-mode DC input (due to $Z_e$) to the amplifier. Keeping a high common-mode rejection coefficient would restrain the DC interference.

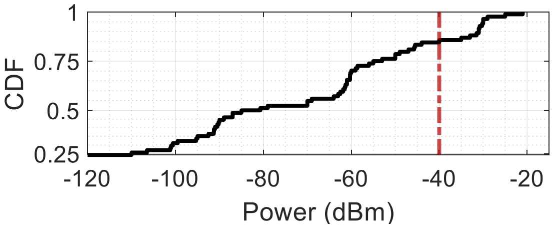


**Figure 6: CDF of the power of $E_{recv}$.** We plug 12 different pairs of earphones into seven different pairing devices and measure the received signal strength at the pairing device in the absence of impedance matching. -40dBm is the minimum power requirement for PCG signal detection.

Figure 5 shows the $E_d$ (received by an oscilloscope) with and without using this voltage detection circuit. Apparently, $E_d$ retains clear S1 and S2 heart sound components after going through this detection circuit, as demonstrated in Figure 5(a). In contrast, as we remove this circuit, we can hardly find the heartbeat cycles on the raw voltage signal receptions, let alone the fine-grained PCG features (Figure 5(b)).

## 3.3 Automatic Impedance Matching (AIM)

The amplified voltage signal $E_d$ flows to the pairing device through the audio chain. Unfortunately, since the impedance of the pairing device $Z_s$ (*i.e.*, the sound card of a laptop) differs from the equivalent impedance of the earphone (*i.e.*, $Z_e$+$Z_p$ in Figure 3), only a small portion of $E_d$ will be absorbed by the pairing device [61], which results in a very weak PCG reception $E_{recv}$ at the pairing device. Our benchmark study shown in Figure 6 further confirms that in most cases the pairing device can hardly receive the PCG signal when we plug the detection circuit directly into the pairing device.[2]

**Programmable impedance matching circuit**. The impedance matching in Asclepius is challenging because both the impedance of earphones $Z_e$ and the pairing device $Z_s$ are unknown in advance. Even worse, their impedance also changes drastically with the hardware type, form factor, and material. To address this issue, we build a programmable impedance circuit using a digital potentiometer chip MAX5402EUA [48]. Its impedance (denoted as $Z_p$) can be programmed with an SPI control signal, which allows us to adapt the earphone's effective impedance ($Z_e$+$Z_p$) to different pairing devices $Z_s$.

**The pitfall in impedance matching**. Conventionally, the impedance matching aims to match $Z_e + Z_p$ to $Z_s$ so that most inductive voltage signal $E_d$ can be delivered to the pairing device (*i.e.*, $E_{recv} \approx E_d$) [61]. However, in Asclepius, as we increase $Z_p$ to match $Z_e + Z_p$ to $Z_s$, the voltage signal $E_d$ will decline (Equation 2), indicating that the sensible PCG signals (represented by $E_d$) become even weaker before arriving at the pairing device. This is particularly detrimental to the higher frequency components (*e.g.*, 100 – 400 Hz) of PCG signals

[2] The impedance of the earphone's speaker is tuned for sound playback, not for sound reception; its impedance mismatches with that of the pairing device.

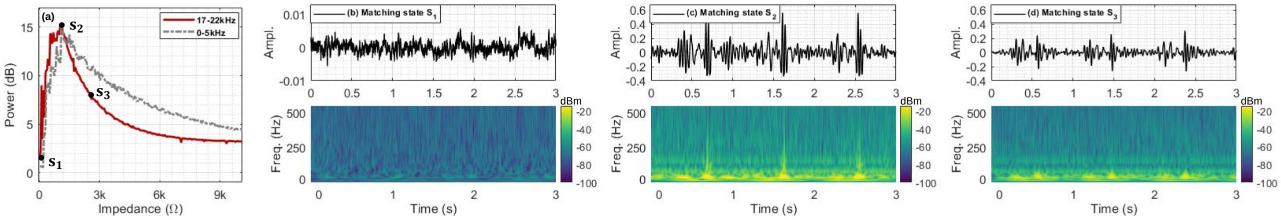


**Figure 7: (a): received signal power in different impedance $Z_p$ settings. (b): the signal profile in the initial, unmatched state ($E_{recv}$ = -62dBm); (c): the signal profile in the optimal, unmatched state ($E_{recv}$ = -25dBm); (d): the signal profile in the fully matched state ($Z_p + Z_e = Z_s$, $E_{recv}$ = -33dBm).**

because these parts are already very weak due to the fact that the higher frequency signals suffer more attenuation when propagating through the human body [35]. Hence adopting the conventional impedance matching principle (*i.e.*, $Z_p$+$Z_e$ = $Z_s$) may not necessarily lead to a better PCG reception.

To validate this argument, we measure the power of the received PCG signal $E_{recv}$ at different impedance settings. As shown in Figure 7(a), $E_{recv}$ grows first and then declines as we increase the impedance $Z_p$. As expected, $E_{recv}$ at the fully matched state $s_3$ is 8dB lower than the signal received at the optimal, unmatched state $s_2$. Moreover, as shown in Figure 7(d), the high-frequency components of PCG signals are overwhelmed by the noise at the fully matched state $s_3$.

**An online impedance tuning algorithm**. To address this pitfall, we propose a feedback-loop-based impedance tuning algorithm to find the optimal matching state. The basic idea is to tune the impedance until we find a matching state that leads to the strongest received signal $E_{recv}$ (*i.e.*, with the highest SNR), as formulated below:

$$\arg\max_{Z_p} SNR(E_{recv}) \tag{3}$$

**Expediting the searching**. Taking each heartbeat symbol as the reference $E_{recv}$ to tune the impedance $Z_p$ would take an excessively long delay since the heart rate is barely around 1–2Hz [53]. To expedite the impedance matching, we send an active probing signal with a very short symbol time (*i.e.*, 10$ms$) from the user's earphone speaker on the right-hand side. This probing signal will propagate through the user's head, captured by the left-ear transducer and our detection circuit inherently. By taking this active probing signal as the reference signal, we can iterate through the searching space within 3 seconds and locate the optimal impedance setting. Specifically, the probing signal consists of consecutive chirps on the ultra-sound (17KHz – 22KHz) band to prevent it from $i$) interfering with the heart sound or motion noises, and $ii$) distracting users. The better noise-resilience of chirp signals allows us to send the probing signals at a lower power (40dBA) and thus makes no harm to human safety [64].

Algorithm 1 describes the impedance tuning process. The *ActiveMatching()* function is called to determine the optimal $Z_p$ value. It iterates through each impedance candidate $i\_Z_p$ within the range of 0-10kΩ [3] and measures the power of the received signal $curr\_E_{recv}$ in each impedance setting using the function *CompEnergy()*. The *CompEnergy()* function contains four steps: i) remove the noise of the received signal $S_i$ using a bandpass filter (BPF) with a cutoff frequency at 17k and 22kHz; ii) down-convert $S_i$ to the baseband (*i.e.*, 0–5kHz) and pass it through a lowpass filter (LPF); iii) remove the possible interference (*e.g.*, modulated physiological signal on the chirp symbol or hardware jitter noise) with a convolution function; and iv) compute the power spectral density (PSD). It is worth noting that down-converting $S_i$ to the baseband will result in better signal quality as LPF retains fewer residual noises at the 3dB cutoff frequency [47] compared to a BPF.

**Algorithm 1:** Online impedance matching

```
input  : $Z_p \leftarrow i\_Z_p$; $\{i\_E_{recv}\} \leftarrow \{\}$;
output : Optimal matching status;
Function ActiveMatching():
    for $i\_Z_p \leftarrow 0$ to $MAX$ do
        $curr\_E_{recv} \leftarrow$ CompEnergy($i\_Z_p$);
        $\{i\_E_{recv}\} \leftarrow$curr_E$_{recv}$;
    end
    $opt\_Z_p \leftarrow$maxitem({i_E$_{recv}$});
    return $opt\_Z_p$;
Function CompEnergy($i$):
    capture audio symbol $S_i$;
    $S_i^* \leftarrow$BPF(S$_i$);
    $S_i^{**} \leftarrow$LPF(S$_i^* \cdot f_{tone}$);
    $S_i^+ \leftarrow$Conv(S$_i^{**}$, $template$);
    $i\_E_{recv} \leftarrow$PSD(S$_i^+$);
    return $i\_E_{recv}$;
```

One may ask would the optimal impedance derived from the ultrasound band be still optimal for the audible band where the heartbeat stays? Figure 7(a) the impedance curve on both the ultrasound band and a 0–5kHz audible band. Notably, the pattern of the in-band impedance curve closely mirrors

[3] The impedance of a pairing device's sound card is usually less than 10kΩ [70] and the impedance of earphones is in the range of 8–600Ω [21].

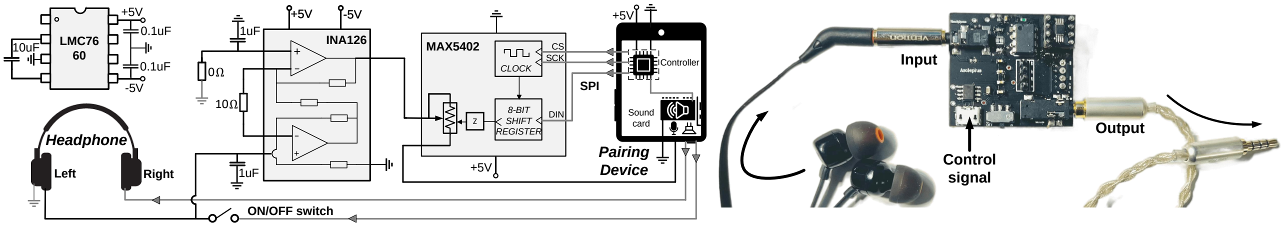


**Figure 8: Schematic (left) and PCB (right) of Asclepius.**

that derived from the 17–22kHz ultrasound probing signal. This similarity can be attributed to the inherent nature of acoustic signals, which makes the impedance less responsive to variations in frequency [82].

## 3.4 Putting Them Together

Figure 8 shows the circuit integration. The schematic contains a low noise amplifier (INA126) for signal detection, a potentiometer chip MAX5402 for automatic impedance matching, and an LMC7660 switched capacitor voltage converter for voltage transformation. The user can turn on/off Asclepius with the onboard switch button. We power this PCB board and send the control signal through a micro-USB interface. The hardware cost is around 5 US dollars.

# 4 ASCLEPIUS'S SOFTWARE

The hardware module adapts the earphone's impedance to the pairing device so that the pairing device can capture the heart sound signals at the cost of a minimum power loss. However, the quality of PCG receptions is low because the PCG signals experience strong attenuation and multi-path effects when propagating inside the human body. Hence the energy and frequency components of PCG receptions will be distorted.

Inspired by the success of deep neural networks (DNN) in signal reconstruction [38, 51, 84], we introduce a data-driven framework to mitigate the frequency and energy distortion in PCG receptions. We envision this framework can be easily integrated into online video visiting platforms as a software patch, serving patients unobtrusively. The overall framework consists of three parts: pre-processing, segmentation, and a two-stage signal recovery. Below we elaborate on each part.

## 4.1 Signal pre-processing

Let $x(t)$ be the PCG signal receptions. The sampling rate of the sound card on the pairing device is set to 48kHz. $x(t)$ undergoes the following three steps.

• **Filtering**. We first filter $x(t)$ with a second-order Butterworth low-pass filter (LPF) with a cutoff frequency at 500Hz to eliminate the out-band noises, *e.g.*, ambient acoustic noises. The cutoff frequency is set based on the fact that the heart sound components such as S1 and S2, as well as murmurs, are in the range of 0 to 500Hz [41, 49, 52].

• **Spike removal**. After filtering, there are still in-band energy spikes that interfere with PCG signals. These energy spikes are due to the friction between earphones and human ears [3, 36]. We then apply a spike removal function to eliminate these energy spikes. Specifically, we divide $x(t)$ into consecutive 500ms time windows with 250ms hop length and compute the maximum absolute amplitudes (MAAs) over each window. If the MAA of a window exceeds the predefined energy threshold (three times the median value of all MAAs), we take it as an outlier spike and remove it from $x(t)$.

• **Normalization**. Finally, we normalize $x(t)$ by scaling it to the range of [-1, 1] and feed the normalized signal into the segmentation step. Such normalization would not affect the fine-grained cardiac characteristics hidden in the collected PCG signals because both the relative amplitude among different heart sound components and their frequencies are well preserved after normalization. Figure 9(a) shows the result.

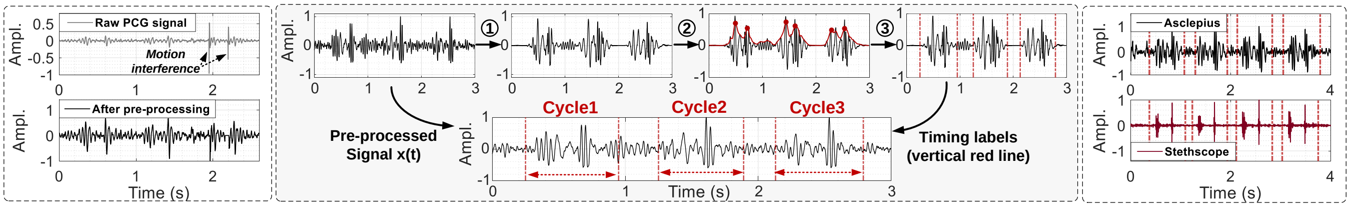
## 4.2 Segmentation

Next, we segment the pre-processed PCG signal $x(t)$ into cardiac cycles [9] for frequency and energy distortion correction. A cardiac cycle describes the sequence of electrical and mechanical events that occurs with every heartbeat. It consists of a heart relaxation (diastole) and a heart contraction (systole) [44]. The duration of a cardiac cycle varies but normally lasts 0.6s – 1s [9]. To ensure the performance of PCG recovery, we have to detect the precise boundary of each cardiac cycle. Below we elaborate on our proposed segmentation method.

• **Signal de-reverberation**. Compared to the clinical PCG signal captured at the human chest, PCG signals captured by earphones propagate over longer distances inside the human body (*i.e.*, from the heart to the ear canal) and thus suffer more from the multi-path effect [35]. These paths have different lengths before reaching the receiver, thus creating different versions that reach at different time intervals. Accordingly, we are expected to see severe reverberations (*i.e.*, inter-symbol interference) on $x(t)$, which makes the boundary of each heartbeat cycle less distinguishable, as shown in Figure 9(b).

Motivated by the success of Wiener filter in ultrasonic imaging de-reverberation [7, 37] and speech enhancement [19, 40, 87], we apply Wiener filter to produce an uncorrupted PCG signal by suppressing the reverberations during diastole intervals [30]. Step 1 in Figure 9 (b) shows the heartbeat signal after applying the Wiener filter. The boundary of each heartbeat cycle after filtering is easily distinguishable.

• **Cardiac cycle segmentation**. Next, we detect the boundary of each cardiac cycle on the de-reverberated PCG signal.



(a) *Signal pre-processing* (b) *Segmentation* (c) *Result*

**Figure 9: Signal preprocessing and segmentation.** (a): The pre-processing removes the in-band interference (*e.g.*, motion artifacts). (b): The segmentation consists of ① signal de-reverberation, ② envelope detector, and ③ cardiac cycle refinement. It detects the precise heartbeat boundary and sends each heartbeat segment to the signal correction and recovery module. (c): The comparison of segmented results with the groundtruth. The heartbeat signals are collected from a healthy 26 years old female.

A straightforward solution would be applying an amplitude threshold to distinguish noise and cardiac signal. However, such a design is susceptible to noise variations and thus is less accurate. In Asclepius, we borrow the hidden Markov model (HMM) based segmentation from biomedical community [29, 72] and propose a *fast boundary detection* and then *refinement* two-phase segmentation method to detect the precise boundary of PCG signals, as explained below.

▹ **Phase One: Fast boundary detection**. We first apply a homomorphic envelope detector [65], followed by a zero-phase low-pass filter [28] to the input (*i.e.*, the de-reverberated PCG signal). The envelope detector keeps the profile of cardiac signals and removes the high-frequency outliers, making S1 and S2 heart sound peaks more prominent (the red curve in Figure 9(b)). Next, we leverage S1 and S2 peaks to detect the coarse-grained boundary of each cardiac cycle using autocorrelation. The span of the cardiac cycle is estimated as the time from lag zero to the highest correlation coefficient.

▹ **Phase Two: Refinement**. The auto-correlation can only detect the averaged length of multiple cardiac cycles. In practice, the length of a cardiac cycle may change over time due to heart rate variability (HRV) [60]. To address this issue, we propose a refinement phase where we search for the precise boundary of each cardiac cycle in the vicinity of the coarse-grained timestamp obtained in the previous step. Specifically, we feed the truncated cardiac cycles into a hidden Marko-based segmentation model (HMM) [72]. The HMM model estimates the probability of the expected precise boundary with logistic regression under the supervision of PCG feature (*e.g.*, S1 and S2 peaks) distributions. In Asclepius, we adopted a public PCG feature distribution. This feature distribution was trained on a large cardiac database [29] and has been proven to be effective in handling both healthy individuals and pathological patients who have bradycardia [4] and tachycardia [79]. The comparison with the ground-truth in Figure 9(c) confirms the efficacy of this design.

## 4.3 PCG signal correction and recovery

The frequency and the phase components of PCG signals are both crucial to auscultations. Motivated by UltraSE [76] in speech enhancement, we propose a two-stage deep learning model (Figure 10) to recover the PCG spectrogram and further refine the PCG waveform in the time domain. The whole process only takes 0.015s to reconstruct a 1.5s heart sound.

- **Stage One: spectrogram recovery**. We adopt a classic encoder-decoder model architecture UNet [63], for PCG spectrogram recovery. UNet has proved its efficacy in human vital sign recovery [14, 32] and signal reconstruction (*e.g.*, magnetic resonance (MR) ) [86]. As shown in Figure 10, the model contains six encoder layers and six decoder layers with skip connections. Each encoder layer consists of a 2D convolution, a batch normalization (BN), a ReLU function, and a dropout regularization module. The stride is set to 2. Each decoder layer comprises a 2D transposed convolution, a BN, a ReLU, and a dropout. Notice that S1 and S2 heart sound components normally last 0.1 second [83]; we thus set the kernel size of the first two convolution layer to 8×8, ensuring its reception field is appropriate to capture a complete S1 and S2 component. Moreover, we replace the standard BN with instance normalization (IN) [80] to expedite training convergence. The frame length of each spectrogram input is set to 2048, with a hope length of 1024. We adopt L1 loss (termed as $L_{spec}$) to measure the difference between the reconstructed PCG spectrogram and the ground-truth spectrogram.
- **Stage Two: waveform refinement**. After the first stage, we will get a PCG spectrogram with reconstructed frequency components. However, the phase values of the reconstructed PCG signals tend to be discontinuous, which will cause inconsistent group delay [13, 31] across frequencies, bringing audible noises to PCG signals. To address this issue, we transform the reconstructed spectrogram to a time-domain waveform using a differentiable iSTFT layer [39] and then propose a second-stage model for waveform refinement.

▹ *Model structure*. We adopt a 1D UNet encoder-decoder model [54] for PCG waveform refinement. Similar to the first-stage model, this 1D UNet also contains six encoder layers and six decoder layers with skip connections. Each encoder layer comprises a 1D convolution, a BN, a PReLU, and a dropout. The PReLU activation function allows the model to accept negative data sample input. The default stride is 2. The decoder layer replaces the convolution with the 1D

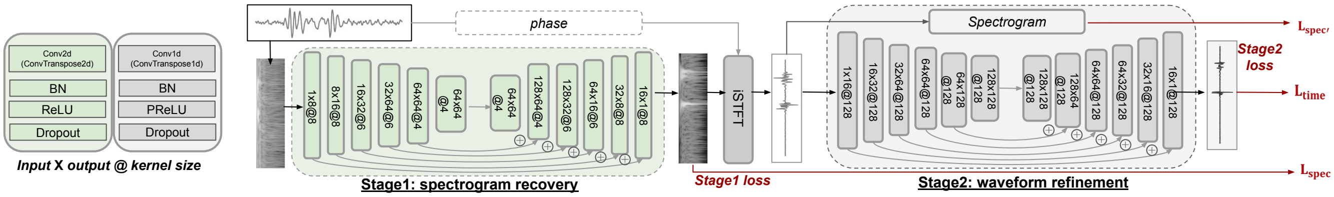


**Figure 10: Two-stage signal recovery model in Asclepius.**

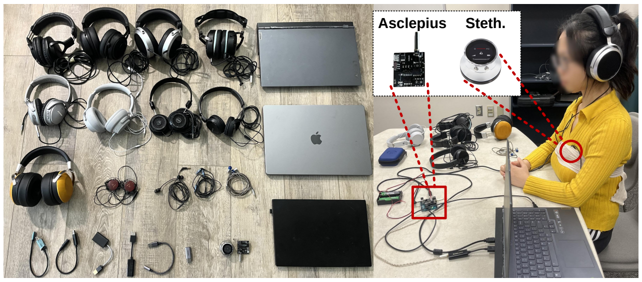


**Figure 11: Earphones and pairing devices used in Asclepius (left); experiment setup on a human subject (right).**

transposed convolution. Note that the audio wave is quasi-stationary within a very short time (2-50 ms) [13], we thus set the kernel size to 128, which ensures a 2 ms reception field on the waveform at 48kHz sampling rate.

▹ *Loss function*. Similar to the stage-one model, we adopt L1 loss to measure the difference between the reconstructed waveform and the ground-truth PCG waveform (termed as $L_{time}$). However, during signal reconstruction, the change of signal samples will alter both the phase and frequency of PCG signals, which may destroy the reconstructed spectrogram. To address this issue, we introduce another L1 loss function $L_{spec'}$ to measure the difference between the reconstructed spectrum after the second stage and the first stage. This loss function will enforce the waveform refinement model to pay attention to phase refinement during waveform reconstruction.

• **Combine two stages together**. These two models are connected in series, and the loss function $L$ is the weighted combination of these three loss functions $L = \alpha * L_{spec} + L_{time} + \beta * L_{spec'}$. The $\alpha$ is manually set to 10 times bigger than $\beta$ to prioritize the spectrogram recovery performance during the training. During the model training, we find the final output PCG waveform contains some high-frequency artifacts above the PCG frequency band occasionally. We thus apply the same second-order low pass filter (§4.1) with 500 Hz cutoff frequencies to the waveform output to eliminate the out-band audio artifacts. The final PCG waveforms are sent to the specialist through the video visit platform.

# 5 EVALUATION

We implement Asclepius's hardware prototype on a 2-layer printed circuit board (PCB). It works as a plug-in peripheral connecting the earphone and the pairing device using 3.5mm audio jacks, as shown in Figure 8. The signal processing pipeline (except for the data-driven PCG signal reconstruction) is implemented in MATLAB. *Due to the page limitation, we put micro-benchmark results and PCG audio samples to an anonymous external link: https://asclepius-system.github.io/*

## 5.1 Experiments Setup

**Data collection**. We collect PCG signals from 30 volunteers (21 males, 9 females) with different ages (22–67 years old), weights, and heights (BMI ranges from 15.9 to 31.8) using different earphones. The ground truth is obtained by an FDA-approved Thinklabs One Digital Stethoscope [74]. The stethoscope is placed at the *Apex* area [68] under the supervision of a medical professional. We set the stethoscope to the *Bell* filter mode [78] to maximize its frequency response for cardiac signal detection while minimizing other physiological sound interference, such as lung sound. The volunteer is asked to keep quiet during the data collection processs to avoid unnecessary motion artifacts, as shown in Figure 11. Each volunteer is asked to fill out a questionnaire for the UX study (§5.5). Overall, 6.7 GB PCG signals are collected.

**Earphone configurations**. The PCG signals are collected by twelve pairs of earphones with different wearing types (over-ear, on-ear, and in-ear), impedance, prices, and transducer sizes. Detailed information about these earphones can be found on our supplementary website. Besides, three different laptops and four different external sound cards are used to capture the PCG signals for further processing.

**Dataset preparation**. We apply the pre-processing algorithm to the raw PCG receptions, segmenting them into heart cycles and zero-padding each heart cycle into 1.5s. Motivated by [32, 67], we adopt leave-one-out cross-validation to evaluate system performance: each time, we train the model on 29 volunteers and test it on another unseen volunteer.

**Model training**. We implement the two-stage signal recovery model on PyTorch 1.6 and train it on a NVIDIA A100 GPU for 200 epochs, with a batch size of 32. We adopt Adam optimizer with a learning rate of 1e-4. We follow a weight-decaying policy at a decaying rate of 90% for every 50 epochs. The hyper-parameter $\alpha$ and $\beta$ are set to 10.0 and 1.0, respectively. We also adopt early stopping to avoid over-fitting.

**Evaluation metric**. Root Mean Squared Error (RMSE) is a widely adopted statistical metric for assessing the quality of PCG de-noising [27, 75] and ECG digitisation [85]. Motivated by them, we adopt the RMSE to quantify the recovered

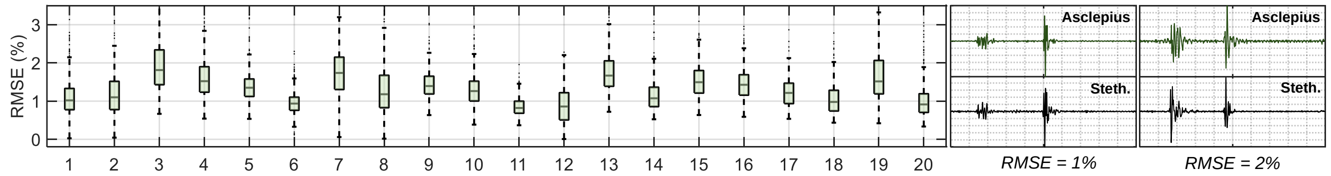


**Figure 12: Overall system performance across 20 selected participants.**

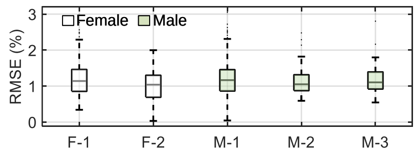


**Figure 13: Different gender and age.**

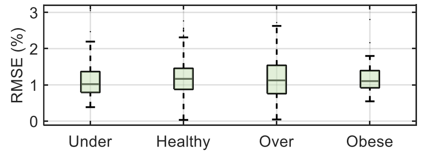


**Figure 14: Different BMI.**

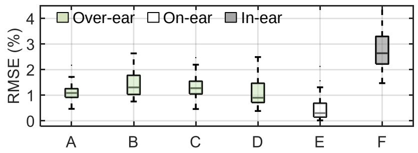


**Figure 15: Different earphones.**

PCG quality in Asclepius. RMSE measures the sample-level difference between the reconstructed PCG and the ground truth using the equation: $RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x(n) - \widetilde{x}(n))^2}$, where $x(n)$ refers to the reconstructed PCG signal; $\widetilde{x}(n)$ refers to the ground-truth PCG samples captured by the stethoscope. Smaller RMSE indicates a higher similarity between the two.

## 5.2 Overall Performance

Figure 12 shows the PCG signal quality of 20 subjects' results randomly chosen from 30 volunteers. Overall, Asclepius achieves decent performance across all 20 participants, with a mean RMSE at 1.34%. For reference, we show the PCG waveform with different RMSE values in the same figure. Taking further scrutiny of these results, we find that subjects 3, 7, 13, and 19 have relatively higher RMSE variances (*e.g.*, >3%) than the remaining subjects. We checked their PCG samples recorded by Asclepius and the stethoscope and find that the PCG signals are partially polluted by noises. This is probably due to unintentional body motions during data collection. We envision a larger training set may help to eliminate the reconstruction bias caused by these motion artifacts. Audio samples can be found at https://asclepius-system.github.io/

- **Impact of age and gender**. Next, we examine the impact of gender and age on PCG signal quality. Restricted by the number of participants, we divide our 30 participants into five groups: F-1 (female, <26 years old), F-2 (female 26–45 years old), M-1 (male, <26 years old), M-2 (male, 26–45 years old), and M-3 (male, >45 years old), respectively. As shown in Figure 13, all five groups achieve consistent PCG signal quality (average RMSE = 1.17%), which indicates that Asclepius is resilient to genders and ages. On the other hand, compared to the group M-2 and M-3, groups F-1, F-2, and M-1 achieve a relatively higher RMSE variance. While we are unsure of the reasons behind this phenomenon, one reason could be that compared to groups F-1, F-2, and M-1, we lack sufficient training samples in groups M-2 and M-3 due to fewer participants. We plan to investigate this issue by recruiting more participants in these two groups.
- **Impact of BMI**. We then examine the impact of different Body Mass Index (BMI) on PCG quality. BMI is a golden-standard measurement of body fat based on the subject's height and weight. We divide 30 participants into four groups, namely, underweight (BMI <18.5), healthy (BMI within 18.5–24.9), overweight (BMI within 25.0–29.9), and obese (BMI >30.0). Figure 14 shows the results. All four groups achieve consistent PCG signal quality (with an average RMSE of 1.09%, 1.19%, 1.13%, 1.24%, respectively), indicating Asclepius is resilient to different BMIs.
- **Impact of earphones**. Next, we evaluate the impact of earphones on the PCG signal quality. In this experiment, we randomly pick one participant from 30 participants and extract the PCG signals collected by six pairs of earphones (out of 12). We then reconstruct these PCG signals with Asclepius and show their signal quality in Figure 15. Overall, we observe that the on-ear earphones achieve the best PCG signal quality (average RMSE = 0.49%), followed by the over-ear earphones (average RMSE =1.22%), and then in-ear earphones (average RMSE = 2.80%). One reason for the superior performance of on-ear earphones is that on-ear earphones have both a large speaker transducer and a short distance to the ear canal. In contrast, although in-ear earphones have even closer contact with the ear canal, their inductive voltage signals due to the heartbeats are relatively weaker due to the smaller size of their speaker transducer. We did not see significant differences in RMSE values of four pairs of over-ear earphones even though their prices vary drastically from 40 to 300 USD.

## 5.3 Comparison study

We further compare Asclepius with HeadFi [24], a state-of-the-art hardware design that reuses speakers on commodity earphones as a microphone to sense physiological activities. To make a fair comparison, we collect PCG signals from seven human subjects using both HeadFi and Asclepius hardware and adopt the same software processing pipeline (§4) introduced in this paper for PCG signal processing and recovery. Worth noting, over 75% pairing devices cannot capture PCG signals with the HeadFi circuit, due to the absence of the

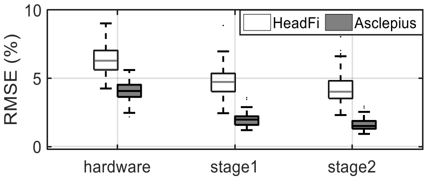


**Figure 16: Comparison study.**

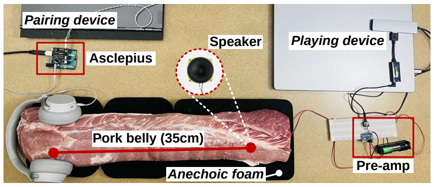


**Figure 17: Experiment setup.**

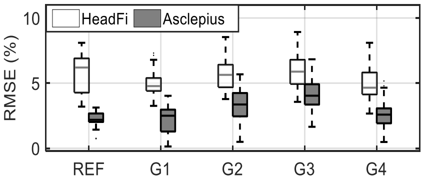


**Figure 18: Emulation results.**

**Table 1: Pathological heart sounds in each group.**

| Group | Explanations |
|---|---|
| **REF** | Normal S1, S2 from a healthy individual. |
| **G1** | Split S1 or S2, absent S2, systolic click, *etc.* |
| **G2** | Holosystolic murmur, early systolic or diastolic murmur, *etc.* |
| **G3** | S3 gallop, S4 gallop. |
| **G4** | Systolic murmur with splitting S2, S3 and holosystolic murmur, *etc.* |

impedance matching design (§3.3). To make HeadFi work, we manually adjusted the circuit impedance of HeadFi, ensuring the successful capture of PCG signals. For our analysis, heart sounds data from four subjects are used as training data, while the data from the remaining three subjects are reserved for evaluation. Both HeadFi and Asclepius data were individually trained, maintaining consistency in signal processing, signal recovery, and hyper-parameter initialization.

**Result**. Figure 16 shows the result. We observe that the average RMSE of the raw PCG signal received by Asclepius's hardware is 3.6%, and the average RMSE of the raw PCG signal received by HeadFi hardware is 6.4% (nearly 2X worse than Asclepius), indicating the effectiveness of Asclepius hardware design. Next, the average RMSE drops to 1.5% as we apply the first-stage signal reconstruction (spectrogram recovery) to Asclepius. In contrast, the average RMSE drops to 4.8% for HeadFi recordings. Furthermore, the average RMSE declines to 0.9% once the second-stage signal reconstruction (waveform refinement) is applied for Asclepius's recording, while the average RMSE maintains 4.3% when the second-stage model is applied for the HeadFi recording. This group of experiments manifests the efficacy of each design component of Asclepius. In the meanwhile, the average RMSE of HeadFi's recording after two-stage improvement (4.3%) is still worse than the raw PCG perception of Asclepius hardware (3.6%). Detailed comparative analysis and audio samples of Asclepius and HeadFi can be found at https://asclepius-system.github.io/

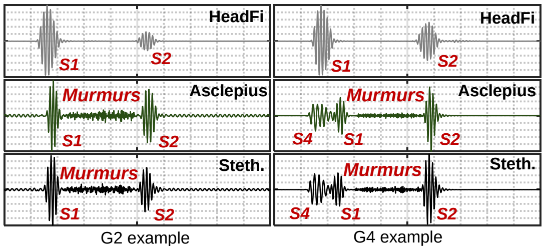
## 5.4 Emulating Patient's Heart Recording

Conducting clinic studies with patients has to undergo a more rigid IRB approval that usually takes more than a year. To examine the efficacy of Asclepius on a patient's heart sound detection, we emulate clinical studies by playing pathological heart sound recordings with a speaker that was placed inside a pork belly. The vibration signals propagate through this pork belly, arriving at the earphones, as shown in Figure 17. These vibration signals undergo multipath fading (*e.g.*, human body) as they travel to the earphone.

**Dataset**. The pathological heart sound recordings are from a public heart sound dataset [81] that was originally used for professional skill training by Umich Medicine. It contains 20 different types of pathological heart sound recordings, each lasting one minute. To emulate different path lengths, we place the speaker in different parts of the pork belly. Moreover, we play the heart sound recordings in different speaker volume settings and the hydration status of the pork belly to emulate human subject variability. In total, We collect 14 hours of PCG signals across 24 different environmental conditions (4 of volume settings × 3 of path length settings × 2 of hydration status). For comparison purposes, an additional 14 hours of PCG signals are collected using HeadFi. Of these, data from 20 conditions were used for training, while the remaining 4 sets were reserved for evaluation.

**Results**. We categorize 20 pathological heart sounds into four groups based on their pathological signal characteristics, namely, G1, G2, G3, and G4. The explanation of each group can be found in Table 1. Additionally, We include a REF group collected from a healthy individual as a reference. The emulation results for both HeadFi and Asclepius are depicted in Figure 18. For Asclepius, we observe the REF group achieves 2.2% RMSE error on average, slightly worse than the results from human subjects-based experiments (§5.2). Upon examining the PCG waveforms, we find that this elevated RMSE stems mainly from the time offset between the captured PCG signal and the ground truth – different from human subject-based experiments where the ground truth and testing data are collected simultaneously and naturally synchronized, the PCG signals in the emulation are collected independently. As a result, we have to align them to the ground truth audio clips manually, which introduces inconsistency. On the other hand, HeadFi's REF group achieves an average RMSE error of 5.8%, which is substantially higher than Asclepius's performance. Furthermore, the RMSE variance for HeadFi is notably higher. This is due to the fact that HeadFi will naturally cancel the PCG signal received by the left-ear transducer and right-ear transducer, resulting in a low-SNR PCG reception. Theoretical analysis can be found on our website.

Taking scrutiny of pathological PCG groups, we observe Asclepius achieves similar signal quality on G1 and the REF groups (with 2.3% RMSE on average), demonstrating that



G2 example

G4 example

**Figure 19: Example pathological signals recovered by HeadFi, Asclepius, and ground-truth (Steth.).**

Asclepius is capable to detect heart diseases specified in this group. Asclepius achieves an average RMSE of 2.8% for group G2, which is slightly worse than G1 and REF. This is reasonable because murmurs in group G2 are high-frequency components that suffer more from attenuation and multi-path effects. We also find large RMSE variations between diastolic and systolic murmurs in this group. Asclepius achieves the worst performance on group G3 (*i.e.*, average RMSE = 4.1%), indicating the detection and reconstruction of S3 and S4 gallop (lower signal amplitude compared to S1 and S2 peaks) is challenging for Asclepius. As for other pathological sound combinations specified in group G4, Asclepius achieves an average RMSE of 2.7%, a comparable performance with REF and G1 group. On the contrary, HeadFi consistently registers high RMSE values (*i.e.*, averaging over 5%) across all pathological groups.

To gain an in-depth understanding of the PCG signal quality provided by both systems, we plot in Figure 19 two representative pathological examples from G2 and G4, each recorded by both HeadFi and Asclepius. The comparison reveals that Asclepius (second row) effectively captures and preserves critical cardiac features, including S1 and S2, high-frequency murmurs, and even the S4 sounds, with its software-hardware co-design. Conversely, HeadFi (first row) predominantly emphasizes the S1 and S2 heart sounds after the signal recovery, while other vital pathological heart features are largely absent.

## 5.5 UX Studies

To further validate the user acceptance and clinical efficacy of Asclepius, we also run UX studies to get feedback from both 30 experiment participants and five cardiologists. Due to the page limit, we put the user feedback part on our anonymous link and only present the two-stage cardiologists' study here:

• **Stage I: Blind testing (objective evaluation)**. In stage I, we devised a blind testing study wherein cardiologists were asked to listen to 50 distinct heart sound clips and diagnose based on each one. Among them, 25 audios are randomly selected from Asclepius's recording, and the other 25 audios with the same cardiac features are from the stethoscope recording. Each group of audio clips encompasses four different categories of cardiac features, including healthy, pathological S1/S2, heart murmurs, and pathological S3/S4. The audio clips are shuffled before playing and the cardiologists are not aware of the audio sources until the end of the blind testing session.

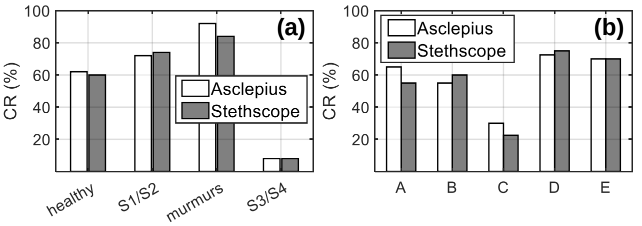


**Figure 20: (a) The percentage of correctly recognized heart sounds over 4 different categories; (b) Accuracy variations among 5 different cardiologists.**

**Result**. Figure 20 (a) shows the percentage of correctly recognized heart sounds in each heart sound category. The result is averaged over five cardiologists' diagnosis results. Overall, we observe that the diagnosis performance based on Asclepius's recordings is quite similar to the performance based on the stethoscope's recordings across all four categories of heart sounds. The diagnosis of murmurs achieves the highest correctness rate (CR) on both Asclepius and stethoscope, followed by the diagnosis of pathological S1/S2 (Asclepius: 72% / stethoscope: 74%). This is because the murmurs are in the high-frequency band and thus are relatively easier to observe compared to S1 and S2 sounds.

Further scrutiny shows that Asclepius achieves a slightly higher diagnosis correctness rate than the stethoscope (92% versus 84%) in detecting the murmurs. One possible reason could be that the stethoscope records may contain more noise interference. In contrast, earphones are less likely to pick up ambient noise when they are put inside the ear canal. The diagnosis correctness rate then drops to 62% (Asclepius) and 60% (stethoscope) for healthy heart sounds. It further declined to a very low level (*i.e.*, 10% for both Asclepius and stethoscope) for pathological S3/S4. The interview with cardiologists revealed that distinguishing S3/S4 sounds is challenging in cardiac auscultation. So cardiologists are less likely to rely on auscultations for diagnosing S3/S4 sounds. They instead focus on S1/S2 and murmurs in auscultation.

Figure 20(b) further shows the diagnosis correctness rate across these five cardiologists. All cardiologists achieve quite similar performance in diagnosing heart sounds from Asclepius's readings and stethoscope's readings. Among five cardiologists, *C* performs slightly worse than the others (*i.e.*, CR <30%). One possible reason could be cardiologist C may not yet have extensive clinical experience and his proficiency in auscultation may be limited.

• **Stage II: Cardiologist interview (subjective evaluation)**. We further designed a UX study under the guidance of a UX researcher and interviewed five cardiologists individually to get their opinions on Asclepius. The interview process was divided into five phases (P1-P5) and hosted online through Zoom. Table 2 shows the dialogue sample from one of the cardiologists. Detailed procedures of interviews and complete dialogues with cardiologists can be found on our website.

**Summary of the interview**. We initiated the interview by briefing the cardiologists about Asclepius. In P1, we asked them what are the most important cardiac features for auscultation. All five cardiologists highlighted the S1 and S2 heart sounds, and heart murmurs as critical features in cardiac auscultation. In P2, cardiologists listened to a healthy individual's PCG signal captured by Asclepius and universally identified the S1 and S2 heart features from the sound. P3 involved a direct comparison between Asclepius's recordings and traditional stethoscope recordings. While the majority found no significant differences, a couple noted slight variations, with descriptions like "less crisp" or "reverberated." P4 expanded on this comparison, focusing on pathological heart sounds. All cardiologists were able to recognize S1, S2, and murmurs sounds in the Asclepius recordings, though one commented on differences in sound intensity. The session wrapped up with a discussion about the strengths and potential limitations of Asclepius. Due to the page limitation, we put all interview results on our website and released one sample in Table 2.

# 6 RELATED WORK

As the next milestone of wearable, earable devices [5, 10–12, 16, 25, 33, 57, 62, 88, 89] have attracted a lot of attention recently. A growing interest in exploring earable techniques [6, 45] is for cardiac monitoring. For example, hEARt [6] utilized an in-ear microphone to monitor heart rate (HR) under both stationary and moving environments. Earmonitor [77] probed FMCW signals to ear canal and captured the ear canal reflections to infer the HR. Similarly, EarACE [8] developed a versatile acoustic sensing platform that is capable to extract PCG envelop and estimating the heart rate variability with the customized ANC earbuds. However, different from Asclepius, these works adopt in-ear microphones for physiological sensing, which are dedicated to costly ANC headphones and are less accessible to the public.

Apart from these adds-on modalities, HeadFi [24], EarSense [59], and other followup [69] explore the speaker transducer on commodity earphones for physiological activity and gesture sensing. However, EarSense achieves this goal by making changes to the soundcard, which is usually prohibited on most PCs and mobiles. HeadFi uses a Wheatstone bridge to remove the music interference. As a side effect, the fine-grained cardiac signals will also be canceled out. Accordingly, it is infeasible to use HeadFi to conduct cardiac auscultation, as we experimentally demonstrated in §5.3. In contrast, Asclepius takes a hardware-software co-design approach to maximize the SNR of the PCG receptions on earphones and further correct frequency distortions of raw PCG receptions due to the multi-path propagation inside the human body, making Asclepius eligible for capturing the detailed S1, S2 heart sounds, as well as the potential heart murmurs.

**Table 2: A sample of the dialog with an anonymous cardiologist. Editing and translation are made for clarity.**

| ▷ **P1: Introducing Asclepius to the clinician:** |
|---|
| **Q1:** *In your clinical experience, what do you consider to be the most crucial aspects of heart sounds when making a diagnosis?*<br>**Answer:** When evaluating heart conditions, it's crucial to carefully assess the primary S1 and S2 heart sounds, as well as any murmurs. While you might also hear S3 and S4 sounds during auscultation, distinguishing between normal and abnormal variants can be challenging. Therefore, the primary focus should always be on the clarity and consistency of the S1 and S2 heart sounds and any identified murmurs. |
| ▷ **P2: Playing PCG signals that Asclepius captured from a healthy individual, and informing the cardiologist that the audio clips are a product of our technology:** |
| **Q2:** *Based on the heart sounds you've just heard, which specific cardiac features can you pinpoint?*<br>**Answer:** I can clearly tell the S1 and S2 components. |
| **Q3:** *How would you compare the heart sounds produced by Asclepius to those you'd typically hear using a stethoscope? Are there any inconsistencies that stood out?*<br>**Answer:** In my experience, I have not observed any discernible differences between the heart sounds produced by your technology and those usually obtained using a stethoscope. The signal quality is exceptional. |
| ▷ **P3: Playing the same PCG signals captured by Asclepius again, then playing the stethoscope recording immediately afterward so the clinician can compare:** |
| **Q4:** *After listening to both, can you tell any differences between the recordings from Asclepius and those from the stethoscope?*<br>**Answer:** Yes, I can tell some differences between these two recordings. Asclepius's recordings are somewhat less crisp compared to those from the stethoscope, and there seem to be some S3 sounds in the background. The stethoscope recordings, on the other hand, have more distinct sounds and no S3 sounds. |
| ▷ **P4: Playing pathological PCG sounds captured by Asclepius. The cardiologist is informed that these clips were produced by our technology and sourced from a patient. After the cardiologist responds to Q5, recordings from the stethoscope are played for comparative analysis:** |
| **Q5:** *Based on these pathological heart sounds you just heard from our system, what cardiac features caught your attention?*<br>**Answer:** I picked up on the S1 and S2 components and also some evident murmurs.<br>**Q6:** *After listening to both, can you pinpoint any differences between Asclepius's recording and the one from the stethoscope?*<br>**Answer:** Honestly, I didn't find any significant differences between the heart sounds from Asclepius and those from the stethoscope. |
| ▷ **P5: Engaging in a conversation with the cardiologist to discuss the advantages and disadvantages of our technology:** |
| **Q7:** *From your expert viewpoint, can you share the benefits you see in using Asclepius?*<br>**Answer:** Certainly. One potential benefit of your technology is that the earphone recording method naturally produces less noise interference compared to a stethoscope. We often face challenges with noise interference when using a stethoscope, which can be caused by factors such as sweat on the skin, environmental noises, and improperly fitted chest contacts. In contrast, earphones are less likely to pick up interference from the ear canal. Additionally, the visual representation of heart sounds in your technology is a significant advantage. We are pleased to have the ability to observe the PCG signal, which will aid in identifying pathological features during auscultation. Furthermore, your system could serve as a valuable tool for remote visits, fostering trust between patients and clinicians by enabling auscultation. |
| **Q8:** *Any thoughts on the limitations and challenges of Asclepius?*<br>**Answer:** A potential challenge I see is tied to the practice of auscultation. Typically, we move the stethoscope to different spots on the chest to obtain better signal quality from specific areas of the heart, such as the right ventricle, pulmonary valve, or tricuspid valve. This allows for an optimized signal quality and comprehensive assessment. With earphones, such precise maneuvering isn't feasible, which might restrict their capacity to capture certain pathological heart activities in these specific areas. |

# 7 CONCLUSION

We have presented the design, implementation, and evaluation of Asclepius, a novel PCG signal detection system using commodity earphones. By listening to the acoustic cardiopulmonary signals captured by Asclepius, the specialists can assess the patient's health condition and make the most informed diagnosis in video visit settings. The evaluation based on 30 participants with various ages and BMI factors confirms the efficacy of Asclepius. The UX studies with these participants and five cardiologists are also positive: over 80% of participants show a willingness to use Asclepius and all cardiologists highly appreciate Asclepius and believe it holds great potential for remote auscultation. Overall Asclepius makes the very first step toward remote auscultation, and we believe it will spark novel ideas in heart sound sensing, pushing the whole field moving forward.

# REFERENCES

[1] C. F. Anderson. Clinical auscultation of the cardiovascular system. *Mayo Clinic Proceedings*, 1990.

[2] Anonymous. Anonymous poster.

[3] E. K. Antonsson, R. W. Mann. The frequency content of gait. *Journal of biomechanics*, 1985.

[4] Bradycardia. https://en.wikipedia.org/wiki/Bradycardia.

[5] N. Bui, N. Pham, J. J. Barnitz, Z. Zou, P. Nguyen, H. Truong, T. Kim, N. Farrow, A. Nguyen, J. Xiao, *et al.* ebp: A wearable system for frequent and comfortable blood pressure monitoring from user's ear. *The 25th annual international conference on mobile computing and networking*, 2019.

[6] K.-J. Butkow, T. Dang, A. Ferlini, D. Ma, C. Mascolo. Motion-resilient heart rate monitoring with in-ear microphones. *arXiv preprint arXiv:2108.09393*, 2021.

[7] N. E. Bylund, M. Ressner, H. Knutsson. 3d wiener filtering to reduce reverberations in ultrasound image sequences. *Scandinavian Conference on Image Analysis*. Springer, 2003.

[8] Y. Cao, C. Cai, A. Yu, F. Li, J. Luo. Earace: Empowering versatile acoustic sensing via earable active noise cancellation platform. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2023.

[9] Cardiac cycle. https://en.wikipedia.org/wiki/Cardiac_cycle.

[10] J. Chan, N. Ali, A. Najafi, A. Meehan, L. R. Mancl, E. Gallagher, R. Bly, S. Gollakota. An off-the-shelf otoacoustic-emission probe for hearing screening via a smartphone. *Nature Biomedical Engineering*, 2022.

[11] J. Chan, A. Glenn, M. Itani, L. R. Mancl, E. Gallagher, R. Bly, S. Patel, S. Gollakota. Wireless earbuds for low-cost hearing screening. *arXiv preprint arXiv:2212.05435*, 2022.

[12] I. Chatterjee, M. Kim, V. Jayaram, S. Gollakota, I. Kemelmacher, S. Patel, S. M. Seitz. Clearbuds: wireless binaural earbuds for learning-based speech enhancement. *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 2022.

[13] T. Chen, L. Shangguan, Z. Li, K. Jamieson. The design and implementation of a steganographic communication system over in-band acoustical channels. *ACM Transactions on Sensor Networks*, 2023.

[14] Z. Chen, T. Zheng, C. Cai, J. Luo. Movi-fi: Motion-robust vital signs waveform recovery via deep interpreted rf sensing. *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021.

[15] F. A. Choudhry, J. T. Grantham, A. T. Rai, J. P. Hogg. Vascular geometry of the extracranial carotid arteries: an analysis of length, diameter, and tortuosity. *Journal of neurointerventional surgery*, 2016.

[16] R. R. Choudhury. Earable computing: A new area to think about. *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, 147–153, 2021.

[17] J. S. Dhillon, C. Ramos, B. C. Wünsche, C. Lutteroth. Designing a web-based telehealth system for elderly people: An interview study in new zealand. *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*, 1–6. IEEE, 2011.

[18] A doctor's touch. https://www.ted.com/talks/abraham_verghese_a_doctor_s_touch.

[19] M. Doerbecker, S. Ernst. Combination of two-channel spectral subtraction and adaptive wiener post-filtering for noise reduction and dereverberation. *1996 8th European Signal Processing Conference (EUSIPCO 1996)*. IEEE, 1996.

[20] J. Eargle. *The Microphone Book: From mono to stereo to surround-a guide to microphone design and application*. Routledge, 2012.

[21] Headphone impedance demystified. https://www.headphonesty.com/2019/04/headphone-impedance-demystified/.

[22] Equivalent series resistance. https://en.wikipedia.org/wiki/Equivalent_series_resistance.

[23] FAITH AND THE STETHOSCOPE. Website.

[24] X. Fan, L. Shangguan, S. Rupavatharam, Y. Zhang, J. Xiong, Y. Ma, R. Howard. Headfi: bringing intelligence to all headphones. *Proceedings of MobiCom*, 2021.

[25] A. Ferlini, D. Ma, R. Harle, C. Mascolo. Eargate: gait-based user identification with in-ear microphones. *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021.

[26] S. N. Gajarawala, J. N. Pelkowski. Telehealth benefits and barriers. *The Journal for Nurse Practitioners*, **17**(2), 218–221, 2021.

[27] S. K. Ghosh, R. K. Tripathy, R. Ponnalagu. Evaluation of performance metrics and denoising of pcg signal using wavelet based decomposition. *IEEE 17th India Council International Conference*, 2020.

[28] D. Gill, N. Gavrieli, N. Intrator. Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model. *Computers in Cardiology, 2005*. IEEE, 2005.

[29] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 2000.

[30] G. Grimmett, D. Stirzaker. *Probability and random processes*. Oxford university press, 2020.

[31] Group delay and phase delay. Website.

[32] U. Ha, S. Assana, F. Adib. Contactless seismocardiography via deep learning radars. *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020.

[33] Y. Jin, Y. Gao, X. Guo, J. Wen, Z. Li, Z. Jin. Earhealth: an earphone-based acoustic otoscope for detection of multiple ear diseases in daily life. *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 2022.

[34] S.-H. Kang, B. Joe, Y. Yoon, G.-Y. Cho, I. Shin, J.-W. Suh, *et al.* Cardiac auscultation using smartphones: pilot study. *JMIR mHealth and uHealth*.

[35] E. Kaniusas. *Acoustical signals of biomechanical systems*. World Scientific, 2007.

[36] R. Khusainov, D. Azzi, I. E. Achumba, S. D. Bersch. Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. *Sensors*, 2013.

[37] K. Kondo, Y. Takahashi, T. Komatsu, T. Nishino, K. Takeda. Computationally efficient single channel dereverberation based on complementary wiener filter. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013.

[38] N. Koonjoo, B. Zhu, G. C. Bagnall, D. Bhutto, M. Rosen. Boosting the signal-to-noise of low-field mri with deep learning image reconstruction. *Scientific reports*, **11**(1), 1–16, 2021.

[39] F. Kreuk, Y. Adi, B. Raj, R. Singh, J. Keshet. Hide and speak: Towards deep neural networks for speech steganography. *arXiv preprint arXiv:1902.03083*, 2019.

[40] F.-J. Kung, M. R. Bai. Estimation of the noise and reverberation covariance matrices with application in speech enhancement using the multichannel wiener filter. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Institute of Noise Control Engineering, 2020.

[41] A. Leatham. *Auscultation of the Heart and Phonocardiography*. Churchill London, 1970.

[42] S. Leng, R. S. Tan, K. T. C. Chai, C. Wang, D. Ghista, L. Zhong. The electronic stethoscope. *Biomedical engineering online*, 2015.

[43] M. Lewkowicz, M. Gitterman. Theory of heart sounds. *Journal of sound and vibration*, **117**(2), 263–275, 1987.

[44] A. A. Luisada, D. M. MacCanon. The phases of the cardiac cycle. *American heart journal*, **83**(5), 705–711, 1972.

[45] D. Ma, A. Ferlini, C. Mascolo. Oesense: employing occlusion effect for in-ear human sensing. *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021.

[46] N. Mamorita, N. Arisaka, R. Isonaka, T. Kawakami, A. Takeuchi. Development of a smartphone app for visualizing heart sounds and murmurs. *Cardiology*, 2017.

[47] M. K. Mandal, S. Sanyal. Compact wideband bandpass filter. *IEEE microwave and wireless components letters*, 2005.

[48] Max5402 datasheet. https://pdfserv.maximintegrated.com/en/ds/MAX5402.pdf.

[49] S. McGee. *Evidence-based physical diagnosis*. Elsevier Health Sciences, 2021.

[50] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sørensen. Transfer characteristics of headphones measured on human ears. *Journal of the Audio Engineering Society*, 1995.

[51] A. Mousavi, A. B. Patel, R. G. Baraniuk. A deep learning approach to structured signal recovery. *2015 53rd annual allerton conference on communication, control, and computing (Allerton)*, 1336–1343. IEEE, 2015.

[52] P. A. Ongley. *Heart sounds and murmurs: A clinical and phonocardiographic study*. Grune & Stratton, 1960.

[53] C. J. Owen, J. P. Wyllie. Determination of heart rate in the baby at birth. *Resuscitation*, 2004.

[54] S. Pascual, A. Bonafonte, J. Serra. Segan: Speech enhancement generative adversarial network. *arXiv preprint arXiv:1703.09452*, 2017.

[55] Patient experience: The clinician connection with patients, matters. https://www.littmann.com/3M/en_US/littmann-stethoscopes/advantages/promotions/clinician-patient-connection/.

[56] A. N. Pelech. The physiology of cardiac auscultation. *Pediatric Clinics*, **51**(6), 1515–1535, 2004.

[57] N. Pham, T. Dinh, Z. Raghebi, T. Kim, N. Bui, P. Nguyen, H. Truong, F. Banaei-Kashani, A. Halbower, T. Dinh, *et al.* Wake: a behind-the-ear wearable system for microsleep detection. *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, 2020.

[58] Report shows overwhelming patient interest in post-pandemic virtual care. Website.

[59] J. Prakash, Z. Yang, Y.-L. Wei, H. Hassanieh, R. R. Choudhury. Earsense: earphones as a teeth activity sensor. *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020.

[60] C. M. van Ravenswaaij-Arts, L. A. Kollee, J. C. Hopman, G. B. Stoelinga, H. P. van Geijn. Heart rate variability. *Annals of internal medicine*, 1993.

[61] Reflection coefficient. https://en.wikipedia.org/wiki/Reflection_coefficient.

[62] T. Röddiger, C. Clarke, P. Breitling, T. Schneegans, H. Zhao, H. Gellersen, M. Beigl. Sensing with earables: A systematic literature review and taxonomy of phenomena. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2022.

[63] O. Ronneberger, P. Fischer, T. Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015.

[64] Safety sound level recommended by epa and who. https://www.cdc.gov/nceh/hearing_loss/what_noises_cause_hearing_loss.html.

[65] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, J. J. Struijk. Segmentation of heart sound recordings by a duration-dependent hidden markov model. *Physiological measurement*, 2010.

[66] V. Schoebel, C. Wayment, M. Gaiser, C. Page, J. Buche, A. J. Beck. Telebehavioral health during the covid-19 pandemic: a qualitative analysis of provider experiences and perspectives. *Telemedicine and e-Health*, **27**(8), 947–954, 2021.

[67] Z. Shi, T. Gu, Y. Zhang, X. Zhang. mmbp: Contact-free millimetre-wave radar based approach to blood pressure measurement. *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, 2022.

[68] N. Silverman, N. Schiller. Apex echocardiography. a two-dimensional technique for evaluating congenital heart disease. *Circulation*, 1978.

[69] X. Song, K. Huang, W. Gao. Facelistener: Recognizing human facial expressions via acoustic sensing on commodity headphones. *Proceedings of IEEE IPSN*, 2022.

[70] Sound cards impedance. https://audioxpress.com/article/practical-test-measurement-sound-cards-for-data-acquisition-in-audio-measurements-part-4.

[71] Can a Speaker be Converted Into an Audio Microphone? Website.

[72] D. B. Springer, L. Tarassenko, G. D. Clifford. Logistic regression-hsmm-based heart sound segmentation. *IEEE transactions on biomedical engineering*, 2015.

[73] Stethoscope. Website.

[74] T. O. D. Stethoscope. https://store.thinklabs.com/products/thinklabs-one-digital-stethoscope.

[75] A. Strazza, A. Sbrollini, M. Olivastrelli, A. Piersanti, S. Tomassini, I. Marcantoni, M. Morettini, S. Fioretti, L. Burattini. Pcg-decompositor: A new method for fetal phonocardiogram filtering based on wavelet transform multi-level decomposition. *Mediterranean Conference on Medical and Biological Engineering and Computing*, 2020.

[76] K. Sun, X. Zhang. Ultrase: single-channel speech enhancement using ultrasound. *Proceedings of the 27th annual international conference on mobile computing and networking*, 160–173, 2021.

[77] X. Sun, J. Xiong, C. Feng, W. Deng, X. Wei, D. Fang, X. Chen. Earmonitor: In-ear motion-resilient acoustic sensing using commodity earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2023.

[78] S. Swarup, A. N. Makaryus. Digital stethoscope: Technology update. *Medical Devices: Evidence and Research*, 2018.

[79] Tachycardia. https://en.wikipedia.org/wiki/Tachycardia.

[80] D. Ulyanov, A. Vedaldi, V. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[81] Umhs michigan heart sound and murmur library. https://www.med.umich.edu/lrc/psb_open/html/repo/primer_heartsound/primer_heartsound.html.

[82] Understanding impedance. https://www.soundonsound.com/techniques/understanding-impedance.

[83] H. K. Walker, W. D. Hall, J. W. Hurst. Clinical methods: the history, physical, and laboratory examinations, 1990.

[84] Y. Wang, D. Wang. A deep neural network for time-domain signal reconstruction. *2015 IEEE International Conference on Acoustics,*

*Speech and Signal Processing (ICASSP)*, 4390–4394. IEEE, 2015.

[85] H. Wu, K. H. K. Patel, X. Li, B. Zhang, C. Galazis, N. Bajaj, A. Sau, X. Shi, L. Sun, Y. Tao, *et al.* A fully-automated paper ecg digitisation algorithm using deep learning. *Nature Scientific Reports*, 2022.

[86] L. Xiang, Y. Chen, W. Chang, Y. Zhan, W. Lin, Q. Wang, D. Shen. Deep-learning-based multi-modal fusion for fast mr reconstruction. *IEEE Transactions on Biomedical Engineering*, 2018.

[87] Z. Xiao, T. Chen, Y. Liu, Z. Li. Mobile phones know your keystrokes through the sounds from finger's tapping on the screen. *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*, 2020.

[88] Z. Yang, R. R. Choudhury. Personalizing head related transfer functions for earables. *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, 2021.

[89] Z. Yang, Y.-L. Wei, S. Shen, R. R. Choudhury. Ear-ar: indoor acoustic augmented reality on earphones. *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020.