Exploring the Feasibility of Remote Cardiac Auscultation Using Earphones

Paper #923: 12 pages, plus references

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 feedback from cardiologists also confirms the efficacy and efficiency of our system. PCG signal samples and benchmark results 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 [15]. 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 expanding their video visit options to reduce healthcare spending, decreasing problems like unnecessary emergency department visits and prolonged hospitalizations [24, 69].

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.

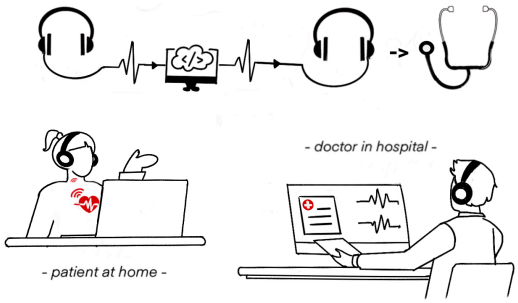

Figure 1: Asclepius empowers the specialist to hear the patient's heart sound during a video visit.

Cardiac auscultation is the most crucial physical examination among different kinds of auscultations [57]. 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 [77] 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 requirement. **High accuracy**. The solution should be able to detect both coarse-grained heart rate variation (HRV) and fine-grained cardiac features (*e.g.*, S1 and S2 sound) 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 human intervention. **Low-cost**. Besides, 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. With these vital sign measurements, the specialist could assess the patient's cardiovascular activities and make objective

therapeutic decisions more precisely. 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 smart/ANC earphones. Hence Asclepius holds great potential to scale to serve large populations.

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) [43], 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 [44]. 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 [18, 22, 61], 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 explained below.

• 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. Finally, the output signal is sent to the specialist for auscultation.

We implement Asclepius's hardware on a 2-layer printed circuit board (PCB). The total hardware cost is around $5 USD. Our software signal processing pipeline is implemented on the pairing device (a laptop) with MATLAB. We evaluate Asclepius using 12 pairs of different commodity earphones. The results based on 30 volunteers of different ages, genders, and BMIs show that Asclepius achieves decent performance – with a 1.17% average Root Mean Squared Error (RMSE) compared to the ground-truth PCG signals. We further emulate clinic testing by playing 20 types of pathological PCG signals using a speaker attached to one end of a pork belly. These vibration signals propagate through the 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 detect pathological PCG signals. Moreover, our UX study shows that over 80% of participants show strong interest in Asclepius, and some of them even take it as a game changer for remote auscultation. The feedback from a cardiologist is also positive – she believes 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 an auxiliary 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 design sketch (§3). We then describe the hardware (§4) and software design (§5). We present the

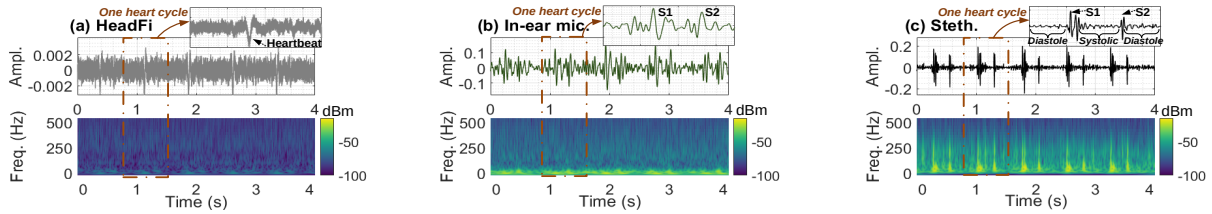

Figure 2: Heartbeat signal of a healthy 28 years old male subject captured by (a) HeadFi, (b) Sound Professional TFB-2 in-ear microphone [89], and (c) a dedicated Thinklabs One digital stethoscope [77].

system evaluation in Section 6. Section 7 describes the related work followed by a conclusion in Section 8.

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 [76] 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 [21]. *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 [16, 56].*

2.2 Remote Auscultation: Opportunities

The demand for video visit services surges during the pandemic and remains strong even after the pandemic [59]. 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 [35, 47] 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 from every possible direction, which may suffer from severe motion artifacts and ambient noises on auscultation.

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

• **Suffer from less ambient noises**. The heart sound propagates through the human body to arrive at the ear canal. The ear cup, ear canal, and eardrum will couple together, forming a hermetic space [51]. 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 [46].

• **Low system cost**. In practice, however, in-ear microphones are only available on those high-end active noise cancellation (ANC) earphones. This inevitably sets a high barrier on the wide adoption of earphones as a stethoscope. Nevertheless, a unique opportunity still exists because every single pair of earphones has two speaker transducers for music playback. Due to the structure reciprocity [74], these speaker transducers can be used as microphones to capture the acoustic signal inside the ear canal. This leaves us with a cost-effective solution for remote auscultation.

• **Easy to operate**. The patient can take remote auscultation in a video clinic visit implicitly and non-intrusively as long as she talks through earphones. Besides, with an earphone, the patient does not need to take the training on placing the stethoscope at the right place.

2.3 Why not HeadFi [22]?

Using speakers on commodity earphones as a microphone to sense physiological activities is not new; both EarSense [61] and HeadFi [22] have already demonstrated such feasibility. However, EarSense requires modifications to the sound card to turn the speaker transducer into a microphone, which is largely prohibited on most sound cards. Although HeadFi has no such requirement, our further scrutiny reveals that HeadFi is not ideal for heart sound monitoring.

Similar to Asclepius, HeadFi is based on an observation that physiological signals such as heartbeats will alter the impedance of the earphone transducer, resulting in an induced voltage signal ΔE. As this induced voltage signal is orders of magnitude weaker than the music playback signal, HeadFi adopts a Wheatstone bridge to cancel out the strong music interference while retaining the subtle voltage signal induced by physiological activities.

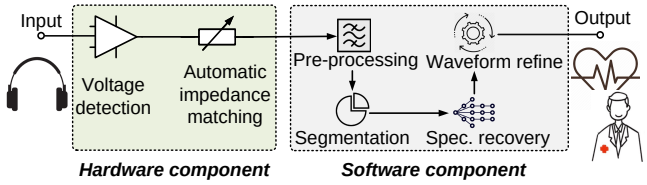

Figure 3: Overview of Asclepius.

The Wheatstone bridge consists of four resistors. Two of them have an identical resistance (impedance), denoted as Z. Another two resistors are replaced by the left transducer and right transducer of an earphone, respectively, which are identical as well. As the physiological signal propagates along different paths before arriving at left and right transducers, it will lead to different variations to the impedance of the left transducer and right transducer of earphones. We denote ΔZ_l and ΔZ_r ($\Delta Z_l \neq \Delta Z_r$) as the impedance variation of the left and right transducer due to physiological activities. The output voltage signal of HeadFi can be represented by:

$$\Delta E \propto \frac{Z \cdot (\Delta Z_l - \Delta Z_r)}{(Z + \Delta Z_r) \cdot (Z + \Delta Z_l)} \tag{1}$$

The above equation reveals that HeadFi captures the difference in ΔZ_l and ΔZ_r, which inherently cancels out fine-grained signal features containing critical physiological meanings. Figure 2 shows the heartbeat signals captured by HeadFi and an FDA-approved digital stethoscope (used as the ground-truth), respectively. We observe that although HeadFi maintains the general profile of each heartbeat cycle, it cannot retain the fine-grained heartbeat features such as S1 and S2 heart sound components that are crucial to auscultation.

3 DESIGN SKETCH

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 through the audio chain. Asclepius explores this opportunity to enable remote cardiac auscultation. Figure 3 shows Asclepius's workflow.

• **Hardware component**. Asclepius's hardware component turns the speaker transducer on headphones into a microphone and delivers the PCG signals to the pairing device (*e.g.*, a desktop or a tablet that the patient uses to talk to the specialist) through the audio chain. We implement this hardware as a low-power plug-in peripheral that wires the patient's earphones to his/her pairing device using two 3.5 mm audio jacks. It takes the cardiac signals as the input, amplifies them using a pre-amp, and tunes the impedance of the earphone with a programmable impedance circuit to ensure most of the cardiac signals can be successfully delivered to the pairing device.

• **Software component**. As the PCG signal suffers from both multi-path effect and attenuation when propagating through the human body, it will experience frequency-selective fading, and the critical frequency components such as murmurs will be distorted. To address this issue, we then propose a data-driven signal-processing pipeline running on the pairing device to recover the fine-grained cardiac features from the raw reception signals. The signal processing pipeline contains four major components, namely, preprocessing, segmentation, spectrogram recovery, and waveform refinement.

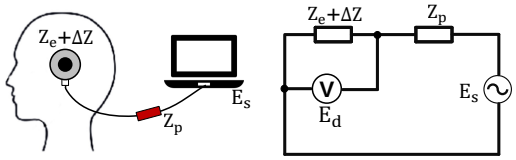

Figure 4: Impedance variation measurement. The change of impedance will alter the voltage E_d.

4 ASCLEPIUS'S HARDWARE DESIGN

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.

4.1 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 4, 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; ΔZ 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:

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

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

$$E_d = \frac{Z_e + \Delta Z}{Z_e + Z_p} \cdot E_s \tag{3}$$

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.

[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.

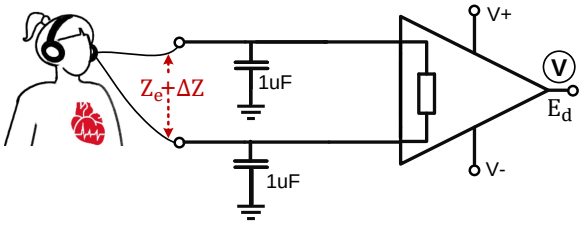

Figure 5: Schematic of the voltage detection circuit.

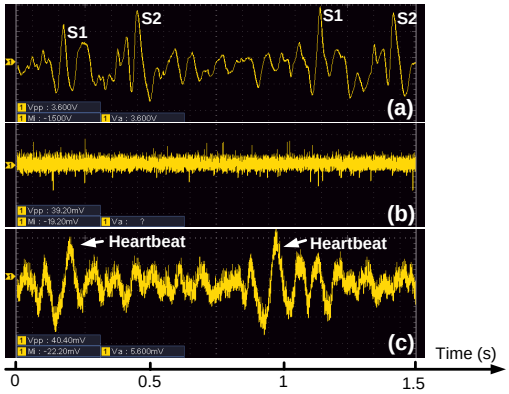

Figure 6: Inductive voltage signal (a) with and (b) without the proposed detection circuit. We also show the signal captured by (c) HeadFi for comparison.

4.2 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 [13]. The right-ear channel is reserved for sound playback.

Figure 5 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 [20] 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 3), not ΔZ 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 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 6(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 6(b)). Our detection circuit also obtains a better signal than HeadFi (Figure 6(c)).

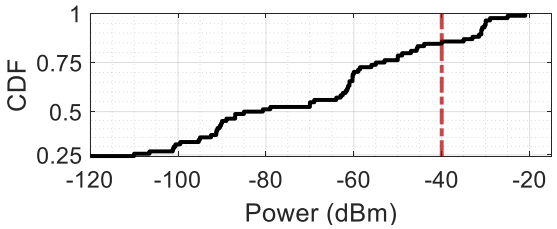

Figure 7: CDF of the power of E_{recv}. We plug 12 different pairs of earphones into seven different pairing devices and measure the received signal strength at the pairing device in the absence of impedance matching. -40dBm is the minimum power requirement for PCG signal detection.

4.3 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 4), only a small portion of E_d will be absorbed by the pairing device [64], which results in a very weak PCG reception E_{recv} at the pairing device. Our benchmark study shown in Figure 7 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. An impedance matching thus is crucial for the successful delivery of PCG signals to 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 [49]. 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. The impedance matching in Asclepius differs from the conventional matching principle. 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.*, $E_{recv} \approx E_d$) [64]. 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 3), 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 because these parts are already

[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.

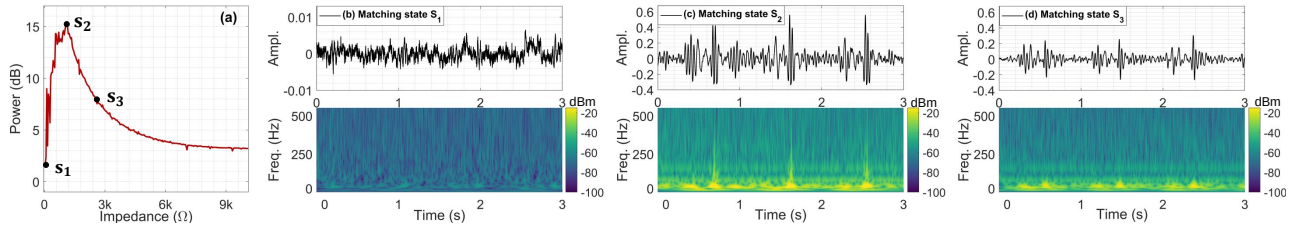

Figure 8: (a): received signal power in different impedance Z_p settings. (b): the signal profile in the initial, unmatched state (E_{recv} = -62dBm); (c): the signal profile in the optimal, unmatched state (E_{recv} = -25dBm); (d): the signal profile in the fully matched state ($Z_p + Z_e = Z_s$, E_{recv} = -33dBm).

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};

very weak due to the fact that the higher frequency signals suffer more attenuation when propagating through the human body [36]. 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 8(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 8(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:

$$\arg\max_{Z_p} SNR(E_{recv}) \tag{4}$$

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 [54]. To expedite the impedance matching, we send an active probing signal with a very short symbol time (*i.e.*, 10ms) 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 [67].

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 [48] compared to a BPF.

[3] The impedance of a pairing device's sound card is usually less than 10kΩ [73] and the impedance of earphones is in the range of 8–600Ω [19].

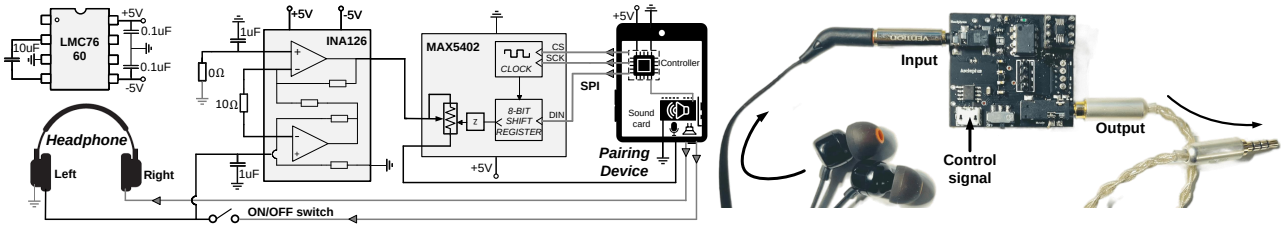

Figure 9: Schematic (left) and PCB (right) of Asclepius.

4.4 Putting Them Together

Figure 9 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.

5 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 [39, 52, 86], 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.

5.1 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 [42, 50, 53].

• **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, 37]. 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 10(a) shows the result.

5.2 Segmentation

Next, we segment the pre-processed PCG signal $x(t)$ into cardiac cycles [8] 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) [45]. The duration of a cardiac cycle varies but normally lasts 0.6s – 1s [8]. 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 [36]. 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 10(b).

Motivated by the success of Wiener filter in ultrasonic imaging de-reverberation [7, 38] and speech enhancement [17, 41], we apply Wiener filter to produce an uncorrupted PCG signal by suppressing the reverberations during diastole intervals [29]. Step 1 in Figure 10 (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 straightforward solution would be applying an amplitude

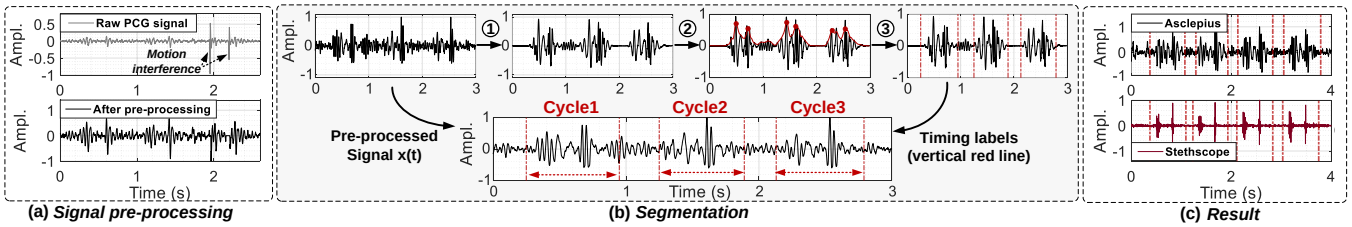

Figure 10: Signal preprocessing and segmentation. (a): The pre-processing removes the in-band interference (*e.g.*, motion artifacts). (b): The segmentation consists of ① signal de-reverberation, ② envelope detector, and ③ cardiac cycle refinement. It detects the precise heartbeat boundary and sends 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.

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 [28, 75] 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 [68], followed by a zero-phase low-pass filter [27] 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 10(b)). Next, we leverage S1 and S2 peaks to detect the coarse-grained boundary of each cardiac cycle using auto-correlation. 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) [63]. 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) [75]. 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 [28] and has been proven to be effective in handling both healthy individuals and pathological patients who have bradycardia [4] and tachycardia [81]. The comparison with the ground-truth in Figure 10(c) confirms the efficacy of this design.

5.3 PCG signal correction and recovery

The frequency and the phase components of PCG signals are both crucial to auscultations. Motivated by UltraSE [79] in speech enhancement, we propose a two-stage deep learning model (Figure 11) to recover the PCG spectrogram and further refine the PCG waveform in the time domain.

• **Stage One: spectrogram recovery**. We adopt a classic encoder-decoder model architecture UNet [66], for PCG spectrogram recovery. UNet has proved its efficacy in human vital sign recovery [12, 31] and signal reconstruction (*e.g.*, magnetic resonance (MR)) [88]. As shown in Figure 11, 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 [84]; 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) [82] 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 [30] 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 [40] and then propose a second-stage model for waveform refinement.

▹ *Model structure*. We adopt a 1D UNet encoder-decoder model [55] 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

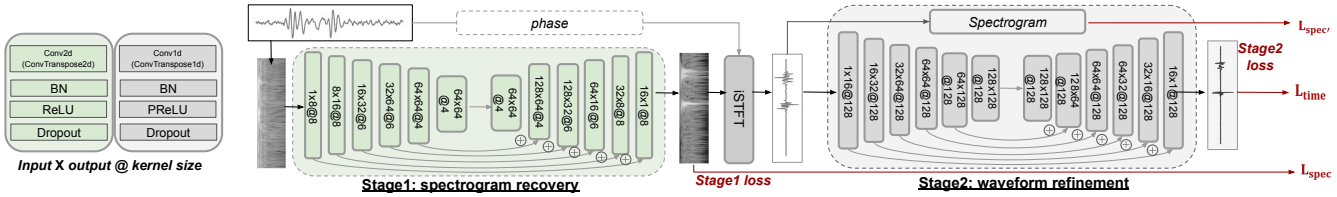

Figure 11: Two-stage signal recovery model in Asclepius.

Figure 12: 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) [91], 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 α is manually set to 10 times bigger than β 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 (§5.1) 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.

6 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 9. 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/*

6.1 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 [77]. The stethoscope is placed at the *Apex* area [71] under the supervision of a medical professional. We set the stethoscope to the *Bell* filter mode [80] 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 12. Each volunteer is asked to fill out a questionnaire for the UX study (§6.4). 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.

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 [31, 70], 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 α and β 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 [26, 78] and ECG digitisation [87]. Motivated by them, we adopt the RMSE to quantify the recovered PCG quality in Asclepius. RMSE measures the sample-level difference between the reconstructed PCG and the ground

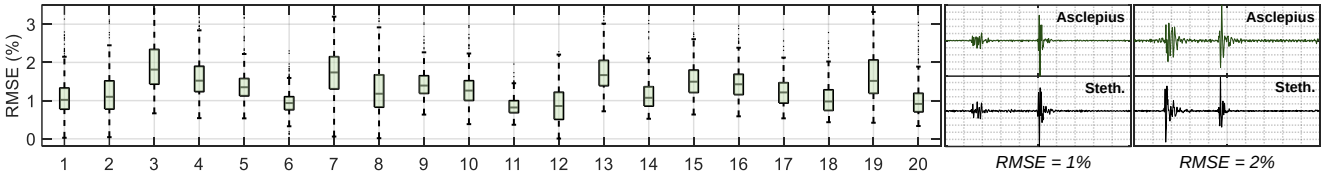

Figure 13: Overall system performance across 20 participants.

Figure 14: Different gender and age.

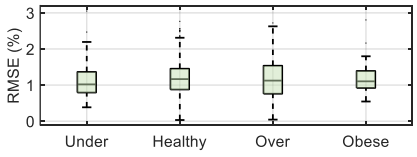

Figure 15: Different BMI.

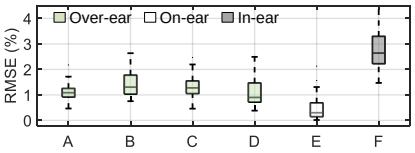

Figure 16: Different earphones.

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.

6.2 Overall Performance

Figure 13 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 14, 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 15 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 16. 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.

• **Ablation study**. Finally, we conduct an ablation study to understand the contribution of each design component to the final output. Figure 17 shows the result. We observe that the average RMSE of the raw PCG signal received by Asclepius's hardware is 3.6%. The average RMSE then drops to 1.5% as we apply the first-stage signal reconstruction (spectrogram recovery). The average RMSE further declines to 0.9% once the second-stage signal reconstruction (waveform refinement) is applied. This group of experiments manifests the efficacy of each design component in Asclepius.

6.3 Emulating Patient's Heart Recording

Conducting clinic studies with patients has to undergo a more rigid IRB approval that usually takes more than half a year. To

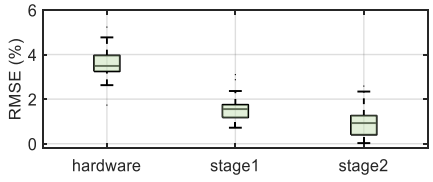

Figure 17: Ablation study.

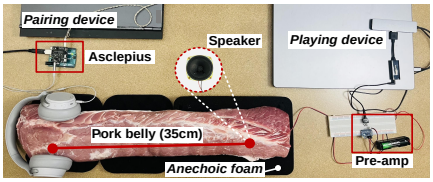

Figure 18: Experiment setup.

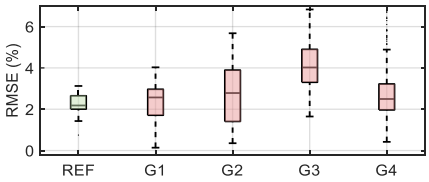

Figure 19: Emulation results.

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.*

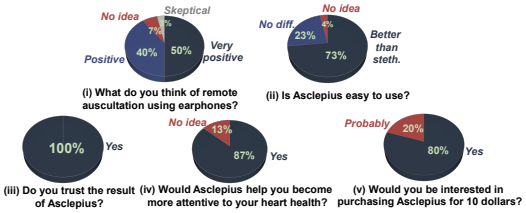

Figure 20: Questionnaire results.

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 18. 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 [83] 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 dryness status of the pork belly to emulate human subject variability. In total, We collect 14 hours of PCG signals under 24 different environmental conditions (4 of volume settings $\times$ 3 of path length settings $\times$ 2 of dryness status). We use data collected in 20 conditions for training and leave the other 4 sets for evaluation.

Results. We divide 20 pathological heart sounds into four groups based on the pathological signal characteristics, namely, G1, G2, G3, and G4. The explanation of each group can be found in Table 1. We also add a REF group collected from a healthy individual as a reference. Figure 19 shows the emulation result. We observe the REF group achieves 2.2% RMSE error on average, slightly worse than the human subjects-based experiments (§6.2). We check their PCG waveforms and find that the large RMSE values for this emulation are primarily due to 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 alone; we have to align them to the ground truth audio clips manually.

Taking a 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 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. We further play these recovered pathological PCG signals to a cardiologist and get very positive feedback (§6.4). Details about Table 1 and audio samples can be found at https://asclepius-system.github.io/

6.4 UX Studies

We also run UX studies to gauge early thoughts around this technology from both 30 experiment participants and cardiologists. Below we describe our findings in UX studies.

UX study I: feedback from users. We design a questionnaire under the guidance of a UX researcher to collect user feedback and comments on remote auscultation and Asclepius. We then ask each participant to fill out the questionnaire before and after the experiment. Figure 20 summarizes the results of a few key questions, out of many. Before the experiment, we briefly explained our technology to each volunteer and asked their opinions. It turns out that 90% of the participants are positive about this technology before trying it. Some participants even commented Asclepius as the game changer of the telemedicine services. 10% of participants either have no idea

about this technology or are skeptical about this technology. The participant then tries our system and is asked to fill out the remaining part of the questionnaire hereafter. The results show that around 73% of volunteers prefer to use Asclepius than a stethoscope for auscultation; 23% of volunteers feel there is no difference between using our system and a stethoscope, and the other 4% said they have "no idea". On the other hand, all volunteers are confident with Asclepius's performance after they see, hear, and further compare their PCG signal recorded by Asclepius and the dedicated stethoscope. Even those 3% skeptical volunteers change their attitude on Asclepius. More importantly, over 87% of participants tell us that Asclepius motivates them to pay more attention to their cardiac health. In terms of the willingness to buy, about 80% of volunteers show interest in purchasing Asclepius at 10 dollars. The remaining 20% said they have a wearable (*e.g.*, an Apple watch or a Fitbit) for health monitoring. But they would not give up the opportunity to use Asclepius.

UX study II: feedback from cardiologists. We interviewed a cardiologist to get her opinion on Asclepius. The interview is divided into four phases (P1–P4). We get approval from the cardiologist and release our dialogue in Table 2.

Table 2: Dialog with a cardiologist. Editing and translation are made for clarity.

▹ ***P1: Introducing Asclepius to the clinician:***
Q1: *What is the most important aspect of heart sounds to consider for diagnosis?* **Answer:** When evaluating heart conditions, it's crucial to carefully assess the primary S1 and S2 heart sounds, as well as any murmurs. Although S3 and S4 sounds may be audible during auscultation, distinguishing between normal and abnormal variants can be challenging. Therefore, the primary focus should be on the S1 and S2 heart sounds, as well as any murmurs identified.
▹ ***P2: Playing PCG signals captured by Asclepius to the cardiologist, and informing her that the audio clips are generated by our technology:***
Q2: *What are your thoughts on the PCG sounds generated by Asclepius? Do you notice any discrepancies compared to what you would expect from a stethoscope?* **Answer:** In my experience, I have not observed any discernible differences between the PCG sounds generated by your technology and those obtained using a stethoscope. The signal quality is exceptional.
▹ ***P3: Playing the PCG signals captured by Asclepius again, then playing the stethoscope recording immediately afterward so the clinician can compare:***
Q3: *Now this time, are you able to differentiate between Asclepius's recording and the stethoscope recording when comparing them side-by-side?* **Answer:** Yes, I can tell some differences in some audio clips between these two recordings. Asclepius's recordings are not as crisp and clear as the stethoscope recordings, and there seems to be some S3 sounds in the background. The stethoscope recordings, on the other hand, have more distinct sounds and no S3 sounds.
Q4: *Would these differences in the signal quality affect your diagnosis in any way?* **Answer:** No, the differences would not affect my diagnosis. Even with actual S3 sounds, it can be difficult to determine whether they are normal or abnormal. Therefore, since I did not detect any changes in the primary S1 and S2 sounds, as well as the murmurs, my diagnosis would remain unchanged.
▹ ***P4: Engaging in a conversation with the cardiologist to discuss the advantages and disadvantages of our technology:***
Q5: *From your perspective, what are the benefits of using Asclepius?* **Answer:** 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.
Q6: *Any thoughts on the limitations and challenges of Asclepius?* **Answer:** One limitation of your technology is that, in real-life auscultation, 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. However, earphones do not allow for this type of maneuver, which could limit their ability to capture certain pathological heart activities that occur in these areas.

7 RELATED WORK

As the next milestone of wearable, earable [14, 65] devices have attracted a lot of attention recently. Examples including exploring earable for improving user experience (*e.g.*, speech enhancement [11, 33] and HRTF [89]), enabling ambient sensing applications such as user authentication [23, 25], human-computer interaction [34, 46, 60, 85], physiological activity monitoring [5, 6, 9, 10, 32, 58, 62], and VR/AR services [90]. However, these system designs heavily rely on dedicated sensors that will add weight, cost, and form factor to earables, setting a strong barrier for their adoption. On the other hand, there is a growing interest in exploring in-ear microphones [6, 23, 46] for physiological sensing. However, in-ear microphones are dedicated to costly ANC headphones and thus are less accessible to the public.

Apart from these adds-on modalities, HeadFi [22], EarSense [61], and other followup [72] 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 §2.3. 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.

8 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 specialist 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 a cardiologist are also positive: over 80% of participants show a willingness to use Asclepius and the cardiologist highly appreciates Asclepius and believes 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] Cardiac cycle. https://en.wikipedia.org/wiki/Cardiac_cycle.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

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

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

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

[21] FAITH AND THE STETHOSCOPE. Website.

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

[23] 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.

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

[25] Y. Gao, W. Wang, V. V. Phoha, W. Sun, Z. Jin. Earecho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2019.

[26] 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.

[27] 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.

[28] 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.

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

[30] Group delay and phase delay. Website.

[31] 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.

[32] 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.

[33] Y. Jin, Y. Gao, X. Xu, S. Choi, J. Li, F. Liu, Z. Li, Z. Jin. Earcommand: " hearing" your silent speech commands in ear. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2022.

[34] V. Kakaraparthi, Q. Shao, C. J. Carver, T. Pham, N. Bui, P. Nguyen, X. Zhou, T. Vu. Facesense: sensing face touch with an ear-worn system. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2021.

[35] 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*.

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

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

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

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

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

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

[46] 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.

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

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

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

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

[51] 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.

[52] 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.

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

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

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

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

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

[58] 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.

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

[60] J. Prakash, Z. Yang, Y.-L. Wei, R. R. Choudhury. Stear: Robust step counting from earables. *Proceedings of the 1st International Workshop on Earable Computing*, 2019.

[61] 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.

[62] M. M. Rahman, T. Ahmed, M. Y. Ahmed, M. Dinh, E. Nemati, J. Kuang, J. A. Gao. Breathebuddy: Tracking real-time breathing exercises for automated biofeedback using commodity earbuds. *Proceedings of the ACM on Human-Computer Interaction*, 2022.

[63] 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.

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

[65] 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.

[66] 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.

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

[68] 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.

[69] 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.

[70] 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.

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

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

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

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

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

[76] Stethoscope. Website.

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

[78] 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.

[79] 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.

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

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

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

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

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

[85] Y. Wang, J. Ding, I. Chatterjee, F. Salemi Parizi, Y. Zhuang, Y. Yan, S. Patel, Y. Shi. Faceori: Tracking head position and orientation using ultrasonic ranging on earphones. *CHI Conference on Human Factors in Computing Systems*, 2022.

[86] 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.

[87] 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.

[88] 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.

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

[90] 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.

[91] E. Zwicker, H. Fastl. *Psychoacoustics: Facts and models*. Springer Science & Business Media, 2013.