EarPCG: Recovering Heart Sounds from in-Ear Audio via Physics-Informed Neural Network

Junyi Zhou[1] Yiyi Zhang[1] Henglin Pu[2] Peng Guo[1] Tianyue Zheng[3] Chao Cai[4] Jun Luo[5]
[1]School of Electronic Information and Communications, Huazhong University of Science and Technology, China
[2]Elmore family School of electrical and computer engineering, Perdue University, USA
[3]School of Computer Science and Engineering, Southern University of Science and Technology, China
[4]College of Life Science and Technology, Huazhong University of Science and Technology, China
[5]College of Computing and Data Science, Nanyang Technological University, Singapore
Email: chriscai@hust.edu.cn, junluo@ntu.edu.sg

Abstract

While earables present a promising avenue for cardiac sensing, whether they may replace the stethoscope to perform heart sound (a.k.a. PCG) monitoring remains questionable. The latest effort attempts to generate PCG-like waveform out of in-ear audio collected via earphones, yet its data-driven approach does not seem to be grounded in the underlying physics. To this end, this paper introduces EarPCG, a system for continuous PCG monitoring leveraging physics-informed neural models. As opposed to the debatable belief that bone-conducted PCG appears within ear canal, EarPCG generates PCG waveforms from the (actually existing) photoplethysmography (PPG) waveforms conveyed via blood vessels. Arising from pressure variations induced by heartbeats, PPG can be mathematically described by a Partial Differential Equation (PDE). Therefore, solving this PDE inversely may reconstruct cardiac dynamics and in turn enable the generation of PCG waveforms with another PDE characterizing the pressure oscillations propagating through soft tissues. Pipelining the two PDE-solving neural models, EarPCG achieves accurate PCG monitoring from in-ear audio, while requiring minimal training. Our extensive experiments leveraging a custom-built prototype demonstrate the efficacy of our proposed system. Furthermore, we have conducted clinical trials, with clinicians reporting no perceptible difference between authentic PCG and the sounds reconstructed by EarPCG.

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing design and evaluation methods**; **Ubiquitous and mobile computing systems and tools**.

Keywords

Earable computing, cardiac monitoring, physics-informed learning.

ACM Reference Format:
J. Zhou, Y. Zhang, H. Pu, P. Guo, T. Zheng, C. Cai, and J. Luo. 2025. EarPCG: Recovering Heart Sounds from in-Ear Audio via Physics-Informed Neural Network. In *The 24th ACM Conference on Embedded Networked Sensor Systems (SenSys'26), May 11-14, 2026, Saint-Malo, France.* ACM, New York, NY, USA, 14 pages. https://doi.org/XXXXXXX.XXXXXXX

* Chao Cai is the corresponding author.

SenSys'26, Saint-Malo, France

ACM ISBN 978-1-4503-XXXX-X/18/06.
https://doi.org/XXXXXXX.XXXXXXX

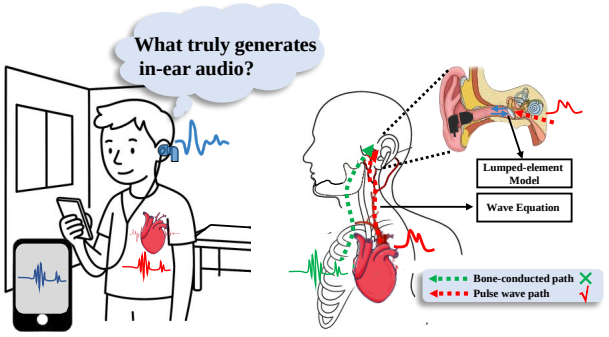


Figure 1: As the existence of PCG in ear canal remains questionable, EarPCG aims to generate PCG from passively collected in-ear audio, grounded in a biophysical model.

1 Introduction

Representing a global health crisis, cardiovascular diseases are responsible for an estimated 17.9 million deaths each year [39]. As the primary tool for practical diagnosis, *stethoscope* allows medical professionals to interpret heart sounds, formally known as the phonocardiogram (PCG). The analysis of PCG irregularities is critical for the early detection of serious conditions such as valve disease [28], congenital heart defects [60], and cardiomyopathy [58]. Despite its diagnostic power, the conventional method of PCG acquisition suffers from significant practical limitations: its reliance on specialized expertise and instrumentation, making it unsuitable for continuous and long-term monitoring outside of a clinical setting [36, 50]. This constraint can delay timely diagnosis and intervention [2, 31, 43], creating a growing demand for alternative PCG monitoring solutions that are continuous and convenient.

In recent years, wireless sensing has emerged as a prominent approach within the research community for *remote* cardiac monitoring [65, 66]. However, these motion-sensing methods face difficulty in "hearing" acoustic PCG remotely as the induced vibration is too weak. Instead, they primarily acquire signal waveforms from body vibrations caused by blood pressure; in other words, wireless sensing obtains only Photoplethysmography (PPG) [1]. For instance, the work in [16] utilizes a wide-band radar to sense miniature chest displacements due to cardiac activity to indirectly retrieve heartbeat waveforms. A similar work from [22] explores mmWave radar

for Seismocardiogram (SCG) sensing, but SCG is not as widely accepted in clinical diagnosis as its acoustic counterpart PCG [69]. Currently, *remote* wireless cardiac monitoring is infeasible due to the notable challenges in distinguishing cardiac activity from respiratory interference [70]. Furthermore, their active sensing nature often entails considerable hardware complexity [16] and high energy consumption. These reasons shift the interest towards current earable sensing.

Earable sensing becomes a promising alternative for those wireless techniques in cardiac monitoring thanks to: (i) the passive sensing nature that offers energy efficiency and reduced hardware complexity; (ii) the occlusion effect that can amplify internal body sounds. This approach is built on the principle that the occluded ear canal passively amplifies faint, bone-conducted signals [35]—a concept successfully applied in areas like user authentication [8, 25] and activity recognition [23, 45]. Inspired by this principle, EarAce [9] claims that PCG transmitted via bone-conduction can be captured with in-ear microphones, and the binaural discrepancy of PCG even carries unique identities [7]. Asclepius [14] further claims to have observed in-ear PCG that suffers from significant frequency and energy loss; it hence involves additional hardware to boost sensitivity and employs a neural network to compensate for these losses. Despite tremendous efforts, this body of work faces a common obstacle: their premise that bone-conducted PCG can be reliably detectable within ear canal lacks rigorous physical validation.

Our measurements, however, challenge the above hypothesis that bone-conducted PCG is directly detectable within the occluded in-ear chamber. Specifically, we observed a consistent temporal latency of approximately 100 ms between the in-ear audio signal and externally measured PCG. This significant delay is inconsistent with the rapid propagation speed of acoustic waves through solid media (i.e., bone) over the relatively short anatomical distances involved, thus questioning the direct bone-conduction pathway as the primary source. Consequently, the direct acquisition of true PCG from within the ear canal via this mechanism appears improbable. This finding led us to propose an alternative mechanism: the in-ear audio is not bone-conducted PCG but is instead acoustic PPG due to arterial pulse wave, propagating through blood vessels connecting the cardiovascular system to the ear. Crucially, this hypothesis perfectly aligns with our key observation. The known propagation velocity of blood pulses (4–12 m/s [46, 64]) precisely accounts for the 100 ms delay we measured. Therefore, while direct PCG acquisition seems unlikely, this robust pulse wave signal provides a new, physically plausible pathway from which to infer cardiac dynamics.

To this end, we introduce EarPCG, a physics-inspired system to infer PCG signals from passive in-ear audio using a comprehensive biophysical model, as shown in Figure 1. EarPCG is a software-oriented solution that (i) requires no extra hardware, (ii) can be deployed on common headsets with in-ear microphones, say the Active Noise Cancellation (ANC) earphones, (iii) and entails even less network parameters than the comparable hardware-software co-optimized approach [14]. Our core innovation is a model of two distinct physical pathways originating from cardiac activity. First, we model the hemodynamic pathway where cardiac pressure waves generate the in-ear audio signal, described by a Partial Differential Equation (PDE) and an Ordinary Differential Equation (ODE). Second, we model the cardiodynamic pathway where the same cardiac activity produces mechanical vibrations that manifest as the PCG. To link these, EarPCG first inverts the hemodynamic model to estimate the source cardiac pressure from the captured in-ear audio. A neural network then transforms this estimated pressure—bridging the gap from hemodynamics to cardiodynamics—into the initial mechanical vibrations, which are propagated through a tissue model to synthesize the final PCG signal. To summarize, this paper makes the following contributions:

- We propose a physical model to analytically explain the root cause of in-ear audio.
- We design a physical-informed deep neural network to accurately reconstruct PCG from in-ear audio.
- We have implemented a system prototype and carried out extensive measurements. Results demonstrate that the proposed model can precisely reconstruct the morphological, diagnostic, and auditory features of PCG. The usability study further confirms high potential for cardiac monitoring.

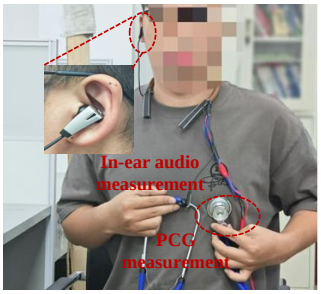
The rest of our paper is organized as follows: Section 2 introduces the background and the motivational measurements, Section 3 presents our system design, and Section 4 demonstrates the implementation of EarPCG. Section 5 reports the experiment results; Section 6 introduces the related work about cardiac status monitoring. Section 7 concludes the paper.

2 BACKGROUND AND MOTIVATION

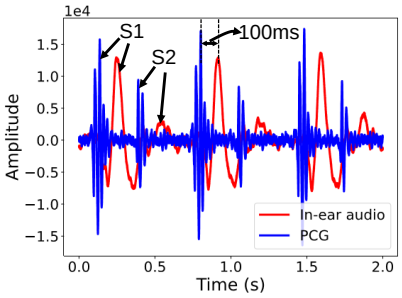
2.1 Background

This section initially presents preliminary studies that explore the use of In-Ear Microphones (IEMs) for cardiac vital sign monitoring. IEMs are hypothesized to capture cardiac physiological signals within the ear canal, an ability attributed primarily to the occlusion effect [35]. The occlusion effect is a phenomenon wherein a confined space, such as the sealed ear canal in this context, passively amplifies low-frequency body sounds through mechanisms like signal superposition and acoustic resonance [10, 56]. Given that the dominant spectral energy of PCG resides within this low-frequency range (20 to 150 Hz [38]) and that in-ear audio captured via IEMs exhibits temporal characteristics resembling PCG, several existing methods commonly treat in-ear audio as bone-conducted PCG.

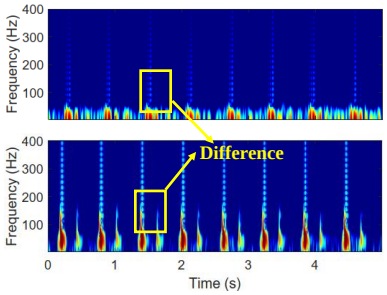
As an example, the authors of [9] engineer a versatile ear-worn sensing platform with IEMs from ANC erables for the extraction of bone-conducted PCG. And they even leverage binaural disparities of in-ear PCGs for user identification [8]. Another work in [6] further utilizes in-ear PCG for Heart Rate (HR) estimation. Zhao et al. [71] utilize the temporal interval between the first and second components of PCG (S1 and S2) identified via in-ear audio, which they also classify as PCG for blood pressure estimation. While the authors of [14] further claim to have captured in-ear PCG that experiences significant frequency and energy loss. They hence design customized hardware with active amplifiers, as well as a sophisticated impedance matching circuit, to increase sensitivity. Meanwhile, a deep neural network is employed to compensate for complex frequency and energy loss. Nonetheless, all these works are built on a premise that bone-conducted PCG can reach the ear canal but without detailed physical validation.



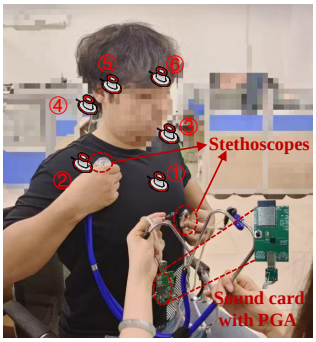
(a) Setup for synchronous data acquisition of in-ear audio and PCG.



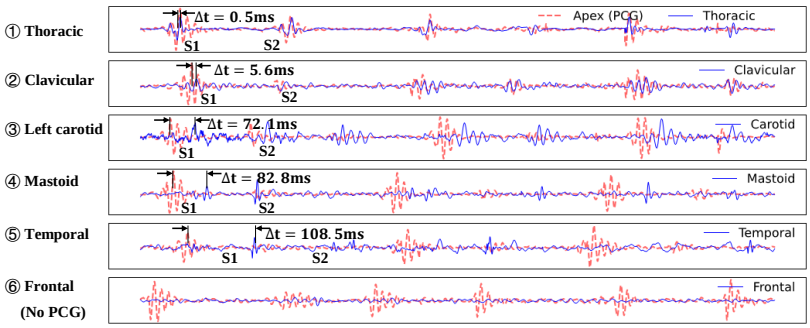
(b) Time domain waveform of PCG and in-ear audio.



(c) Spectrograms of in-ear audio (upper) and PCG (lower).



(d) Setup for synchronized acquisition of PCG in reference and other body locations.



(e) Waveform synchronously acquired at the apex area and other body locations.

Figure 2: Motivational study: (a) The measurement setup of synchronized acquisition of in-ear audio and PCG. (b) PCG and in-ear audio in time domain. (c) Spectrograms of 5 seconds PCG (lower) and corresponding in-ear audio (upper). (d) The measurement setup for synchronized acquisition of PCG in the apex area and the other six locations. (e) The waveform comparison between the (reference) apex area and other body locations.

2.2 Does In-Ear Audio Really Contains Bone-conducted PCG?

To characterize in-ear audio and distinguish it from the PCG, we have conducted a measurement study using our upgraded EarAce platform [12]. We let two synchronous earphones to record PCG and in-ear audio simultaneously. The PCG is recorded by inserting an earphone into the ear tips of a stethoscope, which is positioned on the subject's apex area. Concurrently, a separate earphone is placed in the subject's ear canal to acquire in-ear audio. The setup, as well as results, are shown in Figure 2. Our findings reveal significant temporal and spectral disparities that refute the hypothesis that in-ear audio is simply bone-conducted PCG. First, we observe a consistent latency of approximately 100 ms (calculated by counting the sample index of peaks) in in-ear audio relative to PCG (Figure 2(b)). This significant delay is incompatible with bone conduction; given the respective propagation velocities in bone (4000 m/s) and a stethoscope tube (340 m/s) over similar path lengths, any delay should be negligible (<1 ms)[49]. Second, the signals differ spectrally. While the PCG exhibits a broad frequency range (up to 400 Hz), the in-ear audio's energy is concentrated in the low frequencies (< 80 Hz), as seen in Figure 2(c). A cumulative spectral analysis confirms this, with the 90th percentile of energy at 23 Hz for in-ear audio versus 56 Hz for the PCG. Collectively, these temporal and spectral discrepancies strongly compel us to re-evaluate whether direct PCG acquisition from in-ear canal is feasible.

To verify the above hypothesis, we next carry out more measurement studies. Our setup involves two synchronous stethoscopes: a reference stethoscope at the apex area and a second one at various other body locations, such as the thorax, clavicle, carotid artery, and several points on the head (Figure 2(d)). To maximize sensitivity, we capture the signals with IEMs connected to a Programmable Gain Amplifier (PGA) with 27 dB of gain, a level that would typically clip standard in-ear audio. To more clearly demonstrate the characteristics, the acquired waveforms above head are further amplified. Our results are clear: even though signals resembling PCG at other body locations can be acquired, the temporal latency relative to true PCG refutes that they are bone-conducted PCGs. However, judging from the waveform morphology and spectral features, they are more akin to in-ear audio.

Based on these measurement results, along with our detailed investigation on the anatomical structure of human body, we reasonably suspect in-ear audio may be acoustic PPG signal. PPG signals arise from blood vessel volume changes and can be generated locally throughout the body, making them far more accessible. In contrast, the propagation of PCG signals through the body is fundamentally different, as it can be only sourced by cardiac activities. We hypothesize that PCG signals encounter severe attenuation(more than 45 dB from heart to chest [48]), allowing them to travel only a short distance from the heart. Even if we assume the possibility of PCG reaching the ear bones via bone conduction (which our

data does not support), the signal would still need to traverse soft tissues, causing further damping. More critically, the physics of the occluded ear canal favors PPG (a frequency range from 0 to 20 Hz [15]) detection. The ear canal's small volume (around 2 cc) causes it to function as an acoustic resonator that amplifies very low frequencies (20 Hz) [21]. Consequently, it boosts PPG signals, which fall within this infrasonic range, while distorting or failing to amplify the higher-frequency PCG signals. Therefore, we conclude that it is nearly impossible to directly sense PCG within the ear canal. The primary obstacles are the severe signal attenuation over distance and the overwhelming presence of the PPG signal, which is both locally sourced and amplified by the ear canal itself. Although we attempted to isolate a potential PCG signal with pre-sampling filters, we were still unable to capture it. Crucially, this necessary filtering is absent in existing literature, such as the work by Chen et al. [14]. For these reasons, the signals reported in that study are highly unlikely to be PCG, motivating the needs for further research.

2.3 The Physics that Generates In-ear Audio

Based on our findings, we propose a new model for the generation of in-ear biosignals. We model the sealed ear canal as a closed air cavity acoustically coupled to the nearby carotid artery through soft tissue—a configuration analogous to a cuff-based blood pressure monitor, where a sensor in a sealed chamber indirectly measures arterial pressure waves [21]. Therefore, we hypothesize that the detected signal is not an attenuated PCG, but rather the arterial pulse wave itself, originating from the arteria labyrinthi near the inner ear. This pressure wave causes the tympanic membrane and surrounding tissues to vibrate, which is then detected by the in-ear microphone. This mechanism effectively defines the in-ear signal as a form of acoustic PPG. Crucially, unlike traditional optical counterpart, this acoustic PPG possesses remarkably high fidelity. The proximity of blood vessels to the skin in the ear canal preserves morphological details of cardiac activity that are often lost at other body sites. This high quality is so pronounced that, as we have noted, prior researchers have consistently treated in-ear audio as PCG. This unique combination of accessibility and fidelity makes in-ear PPG an ideal proxy from which to computationally reconstruct true PCG. This allows our system to support more versatile applications than in-ear audio, which loses valuable high-frequency components crucial for diagnosis.

3 System Design

EarPCG is the first passive system that exploits in-ear audio for PCG reconstruction via explicit physical models. This reconstruction is decomposed into three distinct phases, each handled by a corresponding module, as illustrated in Figure 3. The first phase primarily involves a hemodynamic process: a pulse wave, generated by cardiac activity, propagates through blood vessels and then couples to the sealed ear canal. This phase is modeled by a Physics-Informed Neural Networks (PINNs) (HermoNet in Figure 3), which infers the underlying pulse wave dynamics at the heart's origin from in-ear audio. We propose that these dynamics represent part of kinetic energy from cardiac activity. The second phase focuses on the cardiodynamic process, also driven by cardiac activity,

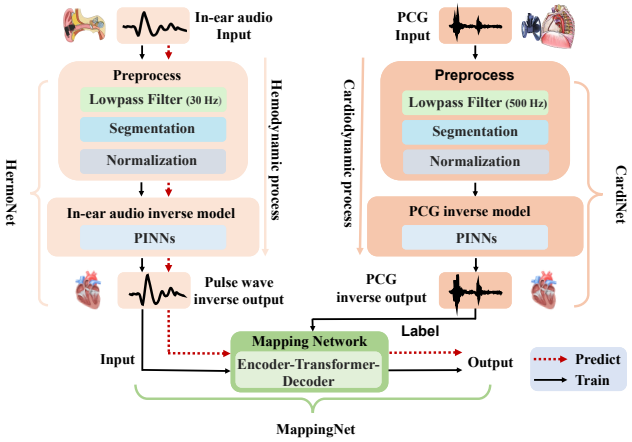


Figure 3: System architecture of EarPCG.

which generates PCG that propagates through soft tissues to the apex area. This process is modeled by another PINNs (CardiNet in Figure 3). While both hemodynamic and cardiodynamic processes originate from the same cardiac activity and are interconnected at the heart. To establish a connection between the aforementioned two processes, the third phase (MappingNet in Figure 3) employs a Transformer network to model this complex relationship.

3.1 Physical Model

In this section, we derive the behind physical models for our EarPCG. To start with, we model the generation of in-ear audio in two stages: first, the propagation of the arterial pulse wave from the heart to the ear; second, the structure-acoustic coupling that converts this pressure wave into a measurable sound in the sealed ear canal, as illustrated in Figure 4.

The first stage, pulse wave propagation (the process 1 in Figure 4), is fundamentally governed by the one-dimensional Navier-Stokes equations for fluid dynamics [30]. These equations relate the arterial pressure wave $p_f(x,t)$ to the vessel cross-section and volume flow rate, where x denotes the distance to the source pressure and t represents time index. While comprehensive, this model contains complex terms for nonlinear convection (inertial effects) and viscous damping (shear stress). However, for blood flow in large arteries, we can introduce two key simplifications. First, under the low Mach number approximation common for arterial flow, the nonlinear convection term can be neglected [42]. Second, viscous effects are primarily confined to the boundary layer and have a minimal impact on the mainstream pulse wave, allowing the damping

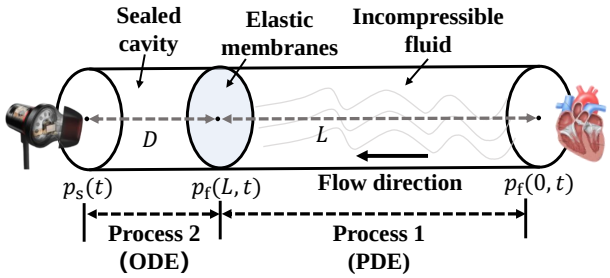


Figure 4: Modeling for the generation of in-ear audio.

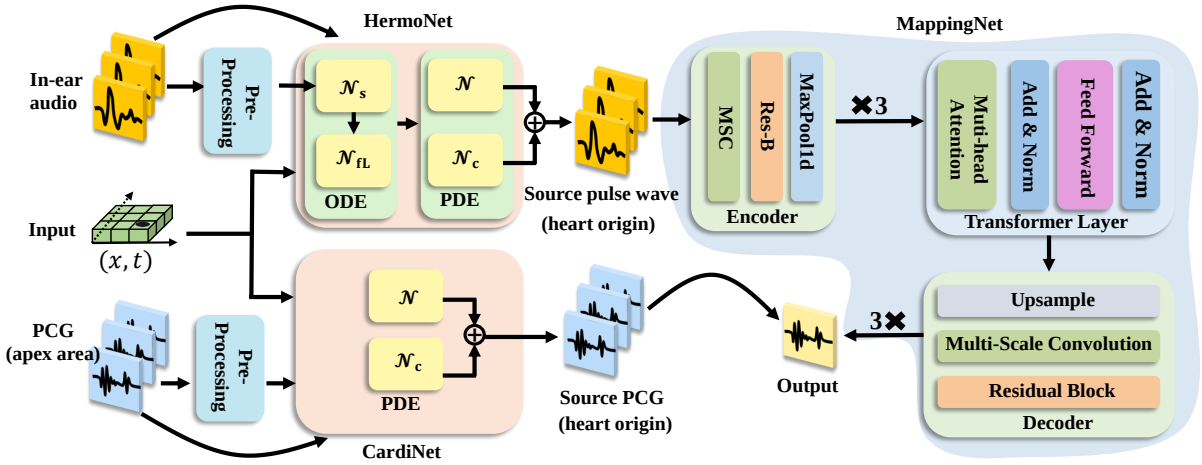


Figure 5: The whole network architecture.

term to be omitted from the one-dimensional averaging theory [19]. After applying these simplifications and introducing the vascular compliance equation [30] to create a closed system of equations, the model reduces to the standard one-dimensional wave equation: $\frac{1}{c^2}\frac{\partial^2 p_f(x,t)}{\partial t^2} = \frac{\partial^2 p_f(x,t)}{\partial x^2}$, where $c = \sqrt{\frac{A}{\rho C}}$ is the pulse wave velocity, which we treat as a learnable parameter constrained to the physiological range of 4–12 m/s. Meanwhile, to account for unmodeled physical phenomena, such as arterial wave reflections or minor nonlinearities, we introduce a compensation term $\Phi(x,t)$, yielding our final propagation model, as shown in Eq. (1).

$$\frac{1}{c^2}\frac{\partial^2 p_f(x,t)}{\partial t^2} = \frac{\partial^2 p_f(x,t)}{\partial x^2} + \Phi(x,t). \tag{1}$$

To ensure model-completeness, the above model's boundary conditions are the aortic pressure wave at the heart origin, $p_f(0,t)$, and the pressure $p_f(L,t)$ arriving near the ear at the end of the vessel.

In the second stage, the arriving pulse wave $p_f(L,t)$ acts on the eardrum and surrounding tissues, generating a sound pressure variation, $p_s(t)$, within the sealed ear canal (process 2 in Figure 4). We model this process by idealizing the eardrum as a forced harmonic oscillator. The equation for the eardrum's displacement, $x(t)$, is given by:

$$m\frac{d^2x(t)}{dt^2} + b\frac{dx(t)}{dt} + kx(t) = A_m(p_f(L,t) - p_s(t)), \tag{2}$$

where m, b, and k are the mass, damping coefficient, and stiffness of the eardrum. A_m is the eardrum's effective area. The key step is to relate the resulting sound pressure $p_s(t)$ to this displacement $x(t)$. By modeling the sealed ear canal as a cavity undergoing isothermal compression, the pressure change can be linearly approximated for small displacements: $p_s(t) = p_t(t) - P_0 \approx P_0\frac{A_m}{V_0}x(t)$, where P_0 and V_0 are the static atmospheric pressure and the canal volume, respectively, and $p_t(t)$ denotes the absolute pressure in the sealed ear canal.

Substituting this linear relationship back into Eq. (2) allows us to eliminate the displacement variable $x(t)$. After rearranging, we arrive at a standard second-order ODE that directly governs the measurable in-ear sound pressure $p_s(t)$ as a response to the arterial pulse wave $p_f(L,t)$:

$$\frac{d^2p_s(t)}{dt^2} + \zeta\frac{dp_s(t)}{dt} + \omega_0^2 p_s(t) = \gamma p_f(L,t). \tag{3}$$

In this final lumped-element model, ζ is the effective damping coefficient, ω_0 is the natural resonant frequency of the coupled eardrum-cavity system, and γ is a coupling gain coefficient that quantifies the efficiency of the pressure transmission. These coefficients, together with those in Eq. (2), reflect subject-specific anatomical variabilities.

The initial phase of our work involves the challenging task of reconstructing the source cardiac pulse waveform at its heart origin based on in-ear sound measurements. This problem can be formulated as identifying the unknown input pressure $p_f(0,t)$ at the heart origin that leads to the observed acoustic pressure $p_s(t)$, but without the usual support of boundary or initial conditions. This lack of prior information classifies it as an ill-posed inversion problem. Moreover, the challenge is further complicated by the presence of unknown parameters $\Phi(x,t)$ within the Eq. (1) and unknown parameter γ in Eq. (3). These unknown conditions render classical numerical methods [40], e.g., finite elements or finite differences, infeasible. Furthermore, the scarcity of labeled data, specifically, pulse waves measured directly at the heart's origin, which are nearly impossible to obtain, precludes the application of traditional deep learning approaches.

In this paper, we employ PINNs [47], a method that combines physical constraints with deep learning models to resolve the first and second phase problem. Compared with the traditional numerical methods, PINNs are mesh-free, without computationally expensive mesh generation, and thus can easily handle the above case [29]. Meanwhile, compared with traditional deep learning, PINNs have the advantages of fast convergence and generalization ability in the small data regime. The embedding of physical information provides the model with prior knowledge, enabling it to make reasonable predictions for unseen data. In particular, the aforementioned subject-specific anatomical variabilities can be readily addressed in its optimization process. The PINNs allow us to resolve both the first and the second phase modeling. For the third phase, we leverage a deep neural network to learn the mapping itself. By incorporating these three modules, we finally built a mapping between in-ear audio $p_s(t)$ to apex area PCG.

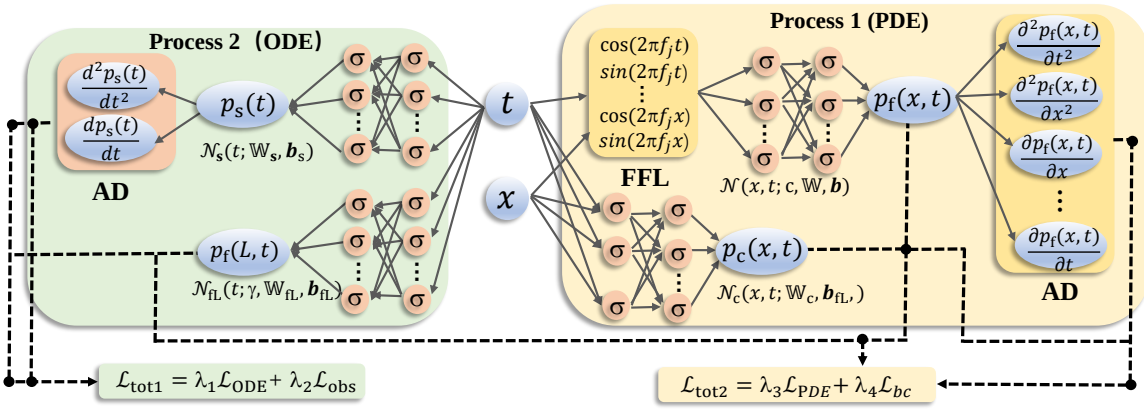


Figure 6: Architecture of HermoNet.

3.2 Networks Design

We resolve the aforementioned hemodynamic and cardiodynamic models using PINNs and establish the mapping relationship leveraging the Transformer. The whole network architecture is illustrated in Figure 5. In this section, we introduce the neural network design.

3.2.1 HermoNet The primary objective of HermoNet is to infer the source pressure waveform $p_f(0,t)$ at the heart's origin based on the observed in-ear audio $p_s(t)$. The problem can be regarded as a typical spatiotemporal inverse problem driven by the coupling of PDE and ODE.

We design the entire network of two parts according to the physical process, as illustrated in Figure 6. The left part of the network (comprising $\mathcal{N}_s$ and $\mathcal{N}_{fL}$) is designed to invert the pressure $p_f(L,t)$ at the distal end, based on the observed in-ear audio $p_s(t)$ (the process 2 in Figure 4), where L denotes the distance from the heart source to the blood vessel near the ear canal. Conversely, the right part (consisting of $\mathcal{N}$ and $\mathcal{N}_c$) further utilizes $p_f(L,t)$ to invert and obtain the pressure waveform $p_f(0,t)$ at the heart (the process 1 in Figure 4). Herein, both $\mathcal{N}_s$ and $\mathcal{N}_{fL}$ are designed using Multiple Layer Perceptron (MLP) with identical architectures, taking t as input. Their outputs correspond to the observed in-ear audio $p_s(t)$ and the inverted distal pressure $p_f(L,t)$, respectively.

In this architecture, the role of $\mathcal{N}_s$ is to fit in-ear audio samples. And $\mathcal{N}_{fL}$ enforces the network to follow the physical laws (defined by Eq. 3). $\mathcal{N}_{fL}$ can be readily implemented via Automatic Differentiation (AD) [4] of a deep learning framework. To train this network, we introduce two loss terms: the observation loss $\mathcal{L}_{obs}$ and the physical loss $\mathcal{L}_{ODE}$. The observation loss L_{obs} is the supervised error term that measures the discrepancy between the neural network's predictions and the observed data; and the physical loss L_{ODE} corresponds to the physical residual loss, an unsupervised term that enforces the governing physics by measuring the discrepancy when the network's output is substituted into the differential equation. Therefore, we define the observation loss $\mathcal{L}_{obs}$ as the squared error between in-ear audio $p_s(t)$ and the output of $\mathcal{N}_s$:

$$\mathcal{L}_{obs} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \|\mathcal{N}_s(t_i; \mathbb{W}_s, \mathbf{b}_s) - p_s(t_i)\|_2, (t \in [0,T]), \quad (4)$$

where $\mathcal{N}_s(t_i; \mathbb{W}_s, \mathbf{b}_s)$ denotes the network function parameterized by $\mathbb{W}_s$ and $\mathbf{b}_s$ and takes time index t_i as its input, $p_s(t_i)$ is the in-ear audio observed at time index t_i, $\|\cdot\|_2$ represents the L_2 norm of the quantities, N_{obs} and T are receptively denote the sample length and duration of the in-ear audio sequence. The loss $\mathcal{L}_{ODE}$ is defined as:

$$\mathcal{L}_{ODE} = \frac{1}{N_{ODE}} \sum_{i=1}^{N_{ODE}} \| \frac{d^2\tilde{p}_s(t_i)}{dt^2} + \zeta \frac{d\tilde{p}_s(t_i)}{dt} + \omega_0^2 \tilde{p}_s(t_i) - \gamma \tilde{p}_f(L,t_i)\|_2, (t \in [0,T]), \quad (5)$$

where $\tilde{p}_s(t_i)$ and $\tilde{p}_f(L,t_i)$ are the predicted outputs of $\mathcal{N}_s$ and $\mathcal{N}_{fL}$ at time index t_i, respectively, N_{ODE} is the number of samples involved in the calculation. Overall, the total loss is formulated as $\mathcal{L}_{tot1} = \lambda_1\mathcal{L}_{ODE} + \lambda_2\mathcal{L}_{obs}$, where λ_1 and λ_2 are two coefficients.

The submodule $\mathcal{N}$ of the network aims to approximate physical process of the pressure dynamics along blood vessels. As shown in Figure 6, the network $\mathcal{N}$ is a MLP that takes coordinates $(x,t), x \in [0,L], t \in [0,T]$ as inputs and outputs estimated pressure waveform $p_f(x,t)$. To train this neural network, we also define two loss terms, a boundary condition loss $\mathcal{L}_{BC}$ and a physics-based loss $\mathcal{L}_{PDE}$. The loss term $\mathcal{L}_{PDE}$ enforces the structure imposed by standard one-dimensional wave equation (Eq. (1)) at a finite set of sampling points; and $\mathcal{L}_{BC}$ guides the model to correct the solution under physical constraints by matching the observed data, improving its ability to approximate the solution. The boundary condition loss $\mathcal{L}_{BC}$ is defined by the square error between predicted pressure waveforms at $x = L$ and the predicted waveform $\tilde{p}_f(L,t)$ from $\mathcal{N}_{fL}$:

$$\mathcal{L}_{BC} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \|\mathcal{N}(L,t_i;c,\mathbb{W},\mathbf{b}) - \tilde{p}_f(L,t_i)\|_2, (x = L, t \in [0,T]), \quad (6)$$

where $\mathcal{N}(L,t_i;\mathbb{W},\mathbf{b})$ denotes the network function. Another loss function $\mathcal{L}_{PDE}$ is constructed by embedding Eq. (1) into the loss:

$$\mathcal{L}_{PDE} = \frac{1}{N_{col}} \sum_{i=1}^{N_{col}} \| \frac{1}{c^2} \frac{\partial^2 \tilde{p}_f(x_i,t_i)}{\partial t^2} - \nabla^2 \tilde{p}_f(x_i,t_i)\|_2, x_i \in [0,L], t_i \in [0,T], \quad (7)$$

where $\{x_i, t_i\}$ are the sampling points (N_{col} in total) in spatial-temporal domain, and $\tilde{p}_f(x_i,t_i)$ is predicted output of network $\mathcal{N}$ corresponding to input $\{x_i,t_i\}$. Each $\frac{\partial^2 \tilde{p}_f}{\partial}$ requires the derivatives of the network output $\tilde{p}_f(x,t)$ with respect to the input (x,t), which are evaluated exactly and efficiently via AD. The total loss is then

constructed by $\mathcal{L}_{\text{tot2}} = \lambda_3 \mathcal{L}_{\text{PDE}} + \lambda_4 \mathcal{L}_{\text{BC}}$, where $\{\lambda_3, \lambda_4\}$ are different weights for each specific loss term. In this paper, an adaptive weight algorithm is applied [61] to address the mismatch in the convergence rate of different losses.

However, due to incomplete physical constraints $\Phi(x,t)$ in Eq. (1), and scarcity of ground truth labels at the heart origin, the above design is hard to converge and tends to exhibit high-frequency oscillations. To this end, we incorporate a Fourier Feature Layer (FFL) [26] and a compensatory network $\mathcal{N}_c$.

The Fourier Feature Layer The inclusion of Fourier Feature Layer (FFL) within the network architecture enables the PINNs to accurately resolve high-frequency, low-amplitude oscillations superimposed on dominant low-frequency signals. This capability significantly improves their overall representational fidelity. The underlying key enabler is that FFL applies a high-dimensional periodic transformation to input coordinates (x,t), which explicitly introduces multiscale frequency information. This transformation allows the network to more effectively capture high-frequency details [26]. In our case, we put FFL at the entrance of $\mathcal{N}$ as shown in Figure 6. This FFL maps the neural network inputs by applying sinusoidal functions (sine and cosine) at multiple frequencies: $F_{\text{out}} = \left[\sin(2\pi f_j x), \cos(2\pi f_j x), \sin(2\pi f_j t), \cos(2\pi f_j t), \dots\right] f_j \in [f_{\min}, f_{\max}]$, where f_j is the j-th frequency in $[f_{\min}, f_{\max}]$.

The Compensation Network The compensation network $\mathcal{N}_c$ here is used to balance the convergence mismatch between boundary loss and PDE loss, due to the remaining unknown term $\Phi(x,t)$ in the governing Eq. (1). Without this module, the boundary loss $\mathcal{L}_{\text{BC}}$ would remain quite high, while the PDE loss $\mathcal{L}_{\text{PDE}}$ is rapidly minimized. Inspired by the work in [72], we introduce a compensation network $\mathcal{N}_c(x,t;\mathbb{W}_c,\mathbf{b}_c)$ to fit the unknown term $\Phi(x,t)$. This network is also represented by an MLP as shown in Figure 6. Its output $p_c(x,t)$ is added to the PDE residual loss to compensate for the mismatch error. Therefore, the term $\mathcal{L}_{\text{PDE}}$ in the loss function Eq. (7) is reformulated as:

$$\mathcal{L}'_{\text{PDE}} = \frac{1}{N_{\text{col}}} \sum_{i=1}^{N_{\text{col}}} \left\| \frac{1}{c^2} \frac{\partial^2 \tilde{p}_f(x_i,t_i)}{\partial t^2} - \nabla^2 \tilde{p}_f(x_i,t_i) - \tilde{p}_c(x_i,t_i) \right\|_2, x_i \in [0,L], t_i \in [0,T]. \tag{8}$$

The total loss $\mathcal{L}_{\text{tot2}}$ is then given by $\mathcal{L}'_{\text{tot2}} = \lambda_3 \mathcal{L}'_{\text{PDE}} + \lambda_4 \mathcal{L}_{\text{BC}}$.

3.2.2 CardiNet The CardiNet module is designed to model the propagation of the PCG from its anatomical source, through soft tissue, to the chest surface. Since PCG is essentially acoustic pressure wave, this tissue-mediated propagation hence can be modeled by a PDE. The primary challenge is that the source PCG is unknown and requires invasive measurement, while the resulting PCG on the chest surface is non-invasively accessible. This scenario—an unknown source and known boundary condition—is an ideal application for PINN. Therefore, CardiNet reuses the PDE solver architecture $(\mathcal{N}, \mathcal{N}_c)$ from HermoNet, but is trained using the non-invasively auscultated PCG as its observation data.

3.2.3 MappingNet The MappingNet is designed to establish a correspondence between pulse waves originating from the heart (obtained in HermoNet) and the reconstructed source PCG (from CardiNet). This establishment would finally allow us to deduce PCG at the apex area from in-ear audio measurements. To formalize this relationship, we train a neural network to learn the nonlinear mapping between them, a simplified schematic of which has been illustrated in Figure 5. To achieve high-fidelity mapping without losing any details in both temporal and spectral features, we adopt a symmetric U-shaped network architecture, specifically an "Encoder-Transformer-Decoder" design. The encoder employs a three-stage progressive structure, integrating Multi-Scale Convolution (MSC) feature fusion, a Residual Block (Res-B), and a Time-Frequency Attention (TFA) module. The Transformer, in particular its self-attention mechanism, has global weight correlations among the multi-scale features extracted by the encoder, enabling the autonomous identification of key temporal patterns in the in-ear audio that are crucial for PCG reconstruction.

The decoder mirrors the encoder structure to maintain a strict spatiotemporal mapping, ensuring that the low-frequency (in-ear audio) features extracted during encoding accurately guide the reconstruction of high-frequency (PCG) details. This symmetric design preserves the temporal dynamics of PCG signals and enhances the reconstruction quality, ensuring that the reconstructed PCG aligns with the rhythmic sense of the PCG obtained on auscultation.

Physic-guided Attention To enhance the model's capability of identifying and reconstructing critical events within the PCG (e.g., S1, S2, premature beats, and murmurs), we introduce a physics-guided attention mechanism, rather than a data-hungry learning method. The key insight of our model is to enforce the network to focus on particular events, for instance S1 and S2, which can be readily and analytically extracted via short-term energy and spectral centroid [37]. This temporal attention, manifested as a binarized time window of resultant events, allows us to apply more weights on the PCG waveform where the critical event occurs, thus preserving its morphology and diagnostic features. This temporal attention is also applied in the time-frequency domain to preserve the spectral features.

Loss function We establish the loss function based on temporal and spectral differences between the source PCG $\mathbf{s}_m$ inferred using the CardioNet and the source pressure wave $\mathbf{s}_p = p_f(0,t)$ predicted by the HermoNet. This composite loss function is formulated as $\mathcal{L} = \alpha M_{\mathbf{TF}} \mathcal{L}_{\text{spec}} + \beta M_{\mathbf{T}} \mathcal{L}_{\text{temp}}$, where α and β are the respective coefficients, $\mathcal{L}_{\text{spec}} = ||STFT(\mathbf{s}_m) - STFT(\mathbf{s}_p)||_2$ denotes the loss computed by the Short Time Fourier Transform (STFT), including both amplitude and phase. And $\mathcal{L}_{\text{temp}} = ||\mathbf{s}_m - \mathbf{s}_p||_2$ characterizes waveform distinctions. The $M_{\mathbf{TF}}$ and $M_{\mathbf{T}}$ represent the time-frequency domain and time domain masks generated by the physics-guided attention mechanism.

4 IMPLEMENTATION

This section elaborates on the implementation details. Our setup uses the upgraded EarAcE platform [12] with a 4 kHz sampling rate. In-ear audio is recorded with W380NB earphones, and the PCG is captured using a YUWELL versatile stethoscope [63]. The neural networks are implemented using PyTorch and trained on NVIDIA 3090 GPUs. We recruit volunteers with a diverse range of ages (21 to 63 years) and Body Mass Index (BMI) values (17.1 to 27.5). To ensure safety, all trials take place on closed test tracks while strictly following our IRB regulations.

4.1 The Network Configurations of EarPCG

The HermoNet architecture is composed of several modules. Its input networks, $\mathcal{N}_s$ and $\mathcal{N}_{fL}$, each consist of two Fully Connected (FC) layers with 64 neurons and a Tanh activation function. The main processing blocks—the backbone network $\mathcal{N}$ and the compensation network $\mathcal{N}_c$—share an identical architecture of five FC layers with 128 neurons each, using a sinusoidal activation function. The crucial distinction is that the input to $\mathcal{N}$ is first encoded by an FFL with a frequency range of 0-20 Hz and a step size of 2 Hz. The CardiNet module adapts this architecture with two key modifications. First, to capture the higher-frequency details of the PCG, its FFL is configured with a wider frequency range (0–200 Hz, 4 Hz step), and its core networks ($\mathcal{N}$ and $\mathcal{N}_c$) are expanded to 256 neurons per layer. Second, CardiNet is structurally simpler, consisting only of the $\mathcal{N}$ and $\mathcal{N}_c$ modules and excluding the $\mathcal{N}_s$ and $\mathcal{N}_{fL}$ networks.

The MappingNet architecture is composed of an encoder, a Transformer module, and a decoder. The encoder consists of three sequential layers, each containing an MSC block, a Res-B, and a max-pooling layer. The MSC block employs four parallel 1D convolutions with varying kernel sizes ([3, 7, 15, 31]), concatenating their outputs before applying Batch Normalization (BN) and PReLU. The Res-B further processes this output, containing another MSC block, a TFA module, a 1D convolution (kernel=3, padding=1), and a final BN. The TFA module is notable for its dual-branch design, using both FFT-based spectral filtering and temporal convolution to generate attention weights. For downsampling, each encoder layer concludes with a max-pooling operation (kernel size=4, stride=4), and a channel-doubling strategy is employed across the layers to increase the feature dimension from 1 to 256. Following the encoder, a Transformer module with four encoder layers processes the sequence. Each layer features an 8-head self-attention mechanism and a Feedforward Network with a dimension of 1024, with a dropout rate of 0.1 applied for regularization. Finally, the decoder reconstructs the signal by first using three up-sampling layers (scale factor=4) to restore temporal resolution, followed by a progressive channel-reducing strategy ([256, 128, 64, 1]) to map the features back to a single-channel output. The whole network has 6.59 M parameters in FP32 format, occupying 25.12 MB memory storage, rendering it plausible to be deployed on constrained mobile devices.

4.2 Pre-Processing and Training

4.2.1 Pre-processing The captured in-ear audio and PCG require pre-processing before being utilized in network training. The pre-processing procedure involves the following stages: 1) low-pass filtering, with a cutoff frequency of 30 Hz for in-ear audio and 500 Hz for PCG, 2) segmentation using an envelope analysis [51], 3) normalization by scaling samples to the range of [-1, 1].

4.2.2 Network Training The training procedure for our system is conducted sequentially, beginning with HemoNet, followed by CardiNet, and concluding with MappingNet. The training of HemoNet proceeds in a two-stage manner, combining both supervised and unsupervised learning. Initially, the input sub-networks, $\mathcal{N}_s$ and $\mathcal{N}_{fL}$, are trained independently for 2000 epochs using the AdamW optimizer with a learning rate of $1e^{-3}$. During this phase, the core PINNs modules ($\mathcal{N}$ and $\mathcal{N}_c$) remain frozen, and network updates are guided solely by the loss term $\mathcal{L}_{tot1}$, with weights $\lambda_1 = 1$ and $\lambda_2 = 50$. Subsequently, the pre-trained input modules are frozen, and training shifts to the core modules $\mathcal{N}$ and $\mathcal{N}_c$. These are optimized according to the loss term $\mathcal{L}_{tot2}$ (with weights $\lambda_3 = 1$, $\lambda_4 = 5$), which comprises a supervised boundary condition loss ($\mathcal{L}_{BC}$) and an unsupervised PDE residual loss ($\mathcal{L}'_{PDE}$). The $\mathcal{L}_{BC}$ is calculated using 4096 boundary points from the $p_f(L, t)$, while the $\mathcal{L}'_{PDE}$ is calculated on 4096 spatiotemporal points sampled via Sobol sequence [41]. This core training is further divided into a 500 epochs pre-training phase using only $\mathcal{L}_{BC}$ with the Adam optimizer (learning rate is $1e^{-3}$), followed by a 2000 epochs fine-tuning phase (learning rate is $8e^{-4}$) where both loss terms are optimized concurrently. An adaptive weighting strategy [61] is also applied every 500 epochs during fine-tuning. The CardiNet follows an identical training protocol for its core modules, with the sole distinction that its boundary condition is defined by the PCG signal acquired from the apex area. Noted that this ground truth PCG is only required at the training phase. During inference, in-ear audio is the only input.

Following the training of the PINNs, the MappingNet is trained in a supervised fashion. It uses the source pressure waveforms from the trained CardiNet as input and the corresponding source PCG signals from the trained CardiNet as the ground-truth. To ensure stable convergence, we employ module-specific initialization strategies: Kaiming initialization for the MSC modules and Xavier initialization for the TFA and Transformer modules. The network is optimized using the AdamW optimizer with a learning rate of $5e^{-4}$, a batch size of 64, and a weight decay of 0.05. The weights for the loss function $\mathcal{L}$ are set to $\alpha = 1$ and $\beta = 10$. To prevent overfitting, we also incorporate an early stopping mechanism.

5 Evaluation

5.1 Feasibility Study

We first conduct a feasibility study to evaluate the ability of our physics-informed model, HemoNet, to accurately infer source signals from distal measurements. Due to the challenges of collecting suitable in-vivo data, we performed this validation using an ex-vivo simulation.

5.1.1 Experiment setup. To validate our physical models with accessible ground truth, we constructed an in-vitro cardiovascular simulation system, as depicted in Figure 7. This system is designed to emulate key biophysical processes: a pulsatile cardiac pump, pressure wave propagation through an arterial pathway, and the resulting pressure fluctuations within a sealed distal cavity that mimics the occluded ear canal. The setup consists of a flow generation unit and a multi-channel measurement system. The flow generator uses a high-precision EDU-P110 peristaltic pump (±1% accuracy) to drive a Gaussian pressure pulse through medical-grade silicone tubing. This tubing terminates in a sealed cavity (length: 23 mm, inner diameter: 6 mm) enclosed by an elastic membrane to simulate the ear canal structure. For data acquisition, a WWL-801M-1M-M20 pressure sensor (0-20 kPa, ±10.5% linearity error) is placed proximally to capture the ground-truth source pressure. A microphone is positioned 40 cm downstream within the sealed cavity to record the distal, observed signal. Both sensors are sampled at 1 kHz with 16-bit resolution via a DAQ122 data acquisition card.

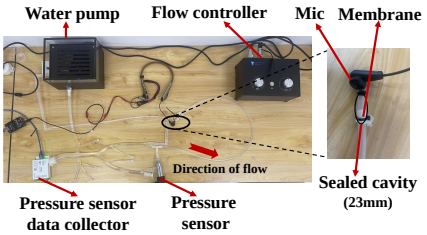


Figure 7: Heartbeat simulation system.

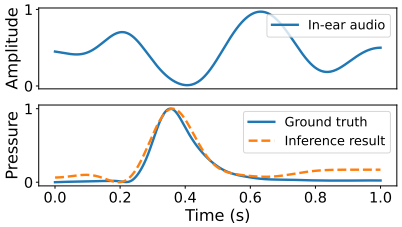


Figure 8: Simulation result.

Figure 9: The setup for data acquisition.

5.1.2 Result Analysis. The reconstructed waveform is shown in Figure 8. To quantify the waveform reconstruction accuracy, we use Root Mean Square Error (RMSE) [14] and Structural Similarity Image Measurement (SSIM) [55]. The RMSE between reconstructed waveform and ground truth is 0.0482, and the SSIM reaches 97.5 %, demonstrating the effectiveness and high accuracy of our model in reconstructing source pressure waveform.

5.2 Overall Performance Evaluation

5.2.1 Experiments Setup To evaluate the performance of our model, we collect data from a cohort of 26 volunteers: 21 healthy individuals and 5 patients with diagnosed heart conditions from Zhongnan Hospital of Wuhan University. The experimental setup is identical to that shown in Figure 2(a). During data collection, participants are instructed to remain still, either sitting or lying down, to minimize motion artifacts (see Figure 9). The audio data are collected under common background noises, including footstep sounds, machine noises, light conversations, etc. To specifically test our model's generalizability under data-scarce conditions, we intentionally create a modest dataset of 1094 audio clips, each lasting for 30 seconds and is sampled at 4kHz. For evaluation, we employ a leave-subject-out cross-validation scheme. Unless otherwise noted, data from one subject is reserved for testing, a second subject's data is used for validation, and the model is trained on the data from the remaining participants.

Evaluation metrics: To assess the clinical utility of our reconstructed waveforms, we move beyond generic signal-level metrics (RMSE, SSIM). And we evaluate four key diagnostic parameters derived from the PCG signal, as illustrated in Figure 10. These include the S1 duration (T_{s1}), S2 duration (T_{s2}), the systolic interval between them (T_{int}), and the S1/S2 energy ratio (E_{ratio}). The formulas for these are given by $T_{s1} = t_{s1e} - t_{s1b}$, $T_{s2} = t_{s2e} - t_{s2b}$, $T_{\text{int}} = t_{s2b} - t_{s1e}$, and $E_{\text{ratio}} = \sum_{t=t_{s1b}}^{t_{s1e}} s(t)^2 / \sum_{t=t_{s2b}}^{t_{s2e}} s(t)^2$ respectively, following established methods [44].

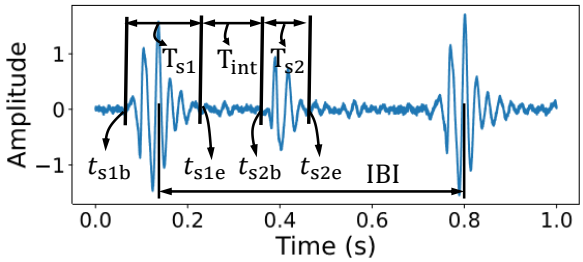


Figure 10: Key physiological metric visualization.

These parameters correspond directly to critical aspects of cardiac function. For example, T_{s1} provides insight into ventricular contraction synchronicity [5], while an extended T_{s2} can indicate conditions like pulmonary hypertension [3]. The systolic interval, T_{int}, is a vital indicator of cardiac functional status [33], and the energy ratio, E_{ratio}, correlates with myocardial contractility and cardiac output [67]. When evaluating the accuracy of these temporal parameters, we establish a clinically relevant error tolerance. Since PCG diagnosis is fundamentally an auditory task, temporal errors below 20 ms are imperceptible to the human ear [13]. Therefore, we consider our model's predictions to be clinically acceptable if the deviation from ground-truth parameters is less than this threshold, as such differences would not affect a physician's diagnosis.

5.2.2 Overall Performance We evaluated the overall performance of EarPCG using a leave-one-out cross-validation methodology across the entire cohort of 26 subjects. Figure 11 illustrates the reconstructed PCG waveforms for these unseen subjects. The system achieves excellent morphological fidelity, with a mean RMSE of 2.935 % and a mean SSIM of 97.148 %. These results demonstrate that EarPCG can accurately reconstruct both the overall shape and fine-grained details of PCG signals.

Beyond visual similarity, we evaluate the model's ability to preserve key physiological signatures. Figure 12 presents the error distributions for four critical diagnostic parameters, augmented with Kernel Density Estimates (KDEs).

For the S1 and S2 durations (Figures 12(a) and 12(b)), the errors are tightly concentrated within ±5 ms. This yields an Mean Absolute Error (MAE) of 3.42 ms for S1 and 3.12 ms for S2, both well below the 10 ms human auditory perception threshold. These correspond to low relative errors of 2.89 % and 3.65 %, respectively. This high precision confirms the model's capability to accurately identify the onset and offset of key cardiac events. Similarly, the error for S1-S2 interval is minimal (Figure 12(c)), with a low MAE of 4.67 ms (2.56 % relative error). This demonstrates the model's accuracy in restoring the systolic timing and overall cardiac rhythm. Lastly, the analysis of the S1/S2 energy ratio (Figure 12(d)) shows that the reconstruction error is tightly controlled within ±0.02. This corresponds to an approximate relative error of 2 %, suggesting that the model effectively preserves the energy characteristics of PCG. In summary, this strong performance indicates that the recovered PCG can be readily used for a range of time-domain cardiac analyses, such as detecting cardiac events and measuring heart rate variability.

5.2.3 Evaluation of Heart Rate Variability We next evaluate the reconstruction of Heart Rate Variability (HRV), a key indicator

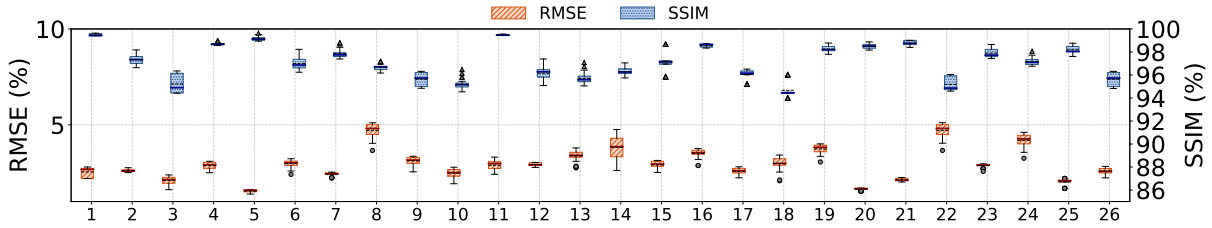
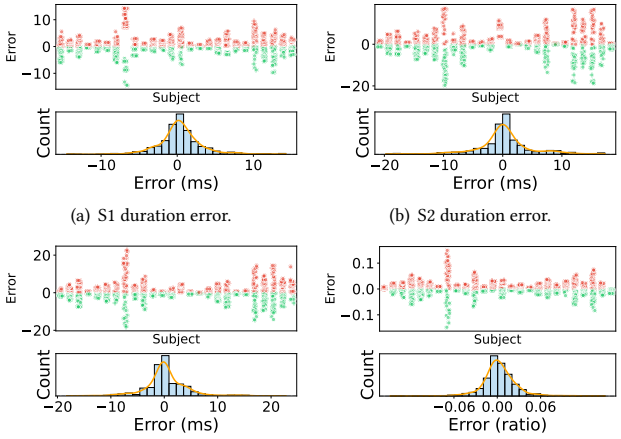


Figure 11: Overall system performance across 26 users.



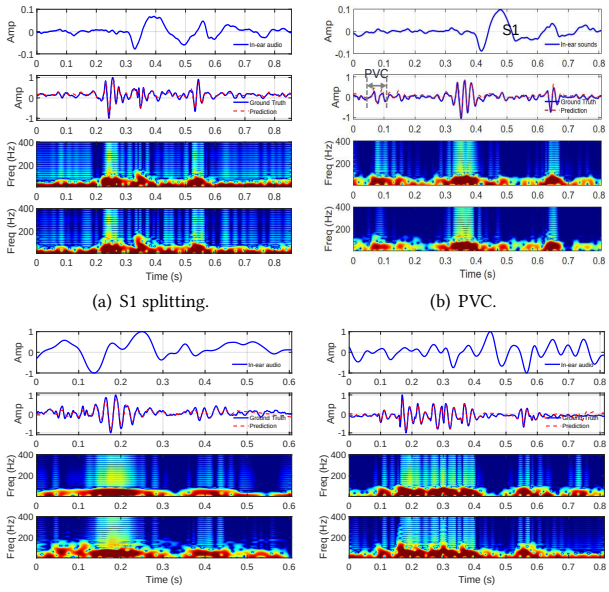
(a) S1 duration error. (b) S2 duration error.

(c) S1/S2 interval error. (d) S1/S2 energy ratio error.

Figure 12: Error distribution of four physiological metrics (S1 duration, S2 duration, S1/S2 interval and S1/S2 energy ratio) related to PCG. For each subplot, the top row shows the distribution of errors for different subjects, while the bottom row presents the corresponding histograms and KDE.

of cardiovascular health derived from beat-to-beat timing fluctuations [54]. We focus on two standard HRV metrics: the Inter-Beat Interval (IBI) and its standard deviation (SDNN). As shown in Figure 14, the reconstructed IBI has a mean error of 6.48 ms, which is only 1 % deviation from the ground-truth average (607.6 ms). The error for SDNN is also low at 6.14 ms, or 4.9 % of the average SDNN (124.6 ms). These low error rates demonstrate that EarPCG accurately captures the dynamic, beat-to-beat fluctuations essential for HRV analysis.

5.2.4 Evaluation of abnormal PCG To further explore its diagnostic capability, we tested EarPCG on subjects with four common cardiac abnormalities: S1 splitting, Premature Ventricular Contraction (PVC) [11], Mitral Regurgitation (MR) [17], and Tricuspid Regurgitation (TR) [53]. These conditions present distinct acoustic signatures: S1 splitting causes a double-peaked S1 sound due to asynchronous ventricular contraction; PVC generates a premature, ectopic beat; and both MR and TR produce continuous systolic murmurs due to incomplete valve closure. Figure 13 presents the reconstructed PCG and their time-frequency spectra. The results demonstrate that EarPCG accurately captures the key pathological features in each case. Quantitative evaluation confirms this high performance. In the time domain, the model achieved high SSIM values (S1 splitting: 93.4 %, PVC: 93.5 %, MR: 91.2 %, TR: 92.6 %) and low RMSE values (0.0726 , 0.0548 , 0.096 , and 0.082 , respectively). This demonstrates excellent waveform fidelity. In the frequency domain, the average



(a) S1 splitting. (b) PVC.

(c) Mitral regurgitation. (d) Mitral and tricuspid regurgitation.

Figure 13: Comparison between the ground truth and reconstructed waveforms and their corresponding spectrograms for abnormal PCG, including (a) S1 splitting, (b) PVC, (c) MR, and (d) MR and TR. From top to bottom in each sub-figure are in-ear audio, time-domain waveforms including reconstructed and ground truth PCG, spectra of reconstructed PCG, and ground truth PCG.

log-spectral distance [32] and spectral convergence [52] across the four conditions were 2.36 dB and 0.084. These low values signify that the auditory difference is negligible, preserving the diagnostic information for clinicians.

5.3 Ablation Studies

Next, we carry out ablation studies to investigate the contributions of physics-inspired modules, including PINNs and physics-guided attention, to performance gain.

5.3.1 Performance With and Without PINNs To quantify the contribution of our physics-informed design, we conduct an ablation study. We compare our full model against a baseline network with an identical architecture but trained without the PINN-based constraints and loss function. The results confirm that incorporating physics provides critical advantages in both training efficiency and generalization.

The benefits are apparent as demonstrated by results shown in Figure 15. The PINN-informed model demonstrates significantly

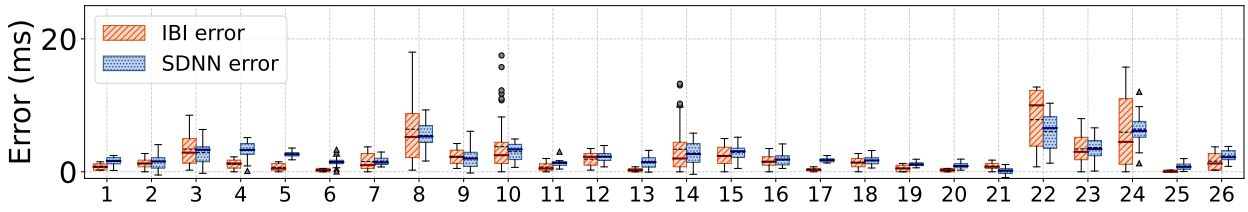


Figure 14: Absolute errors of IBIs and heart beat SDNN.

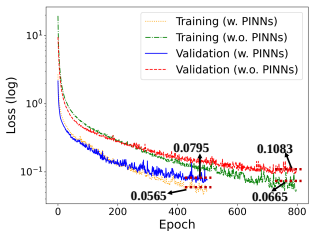


Figure 15: The loss curve w. and w.o. PINNs.

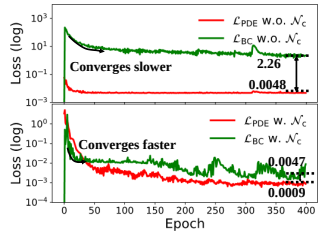


Figure 16: The curve w. and w.o. $\mathcal{N}_c$.

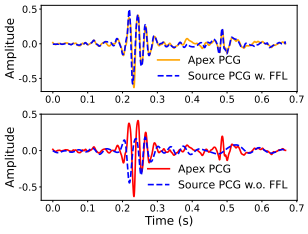


Figure 18: Waveform reconstructed w. and w.o. FFL.

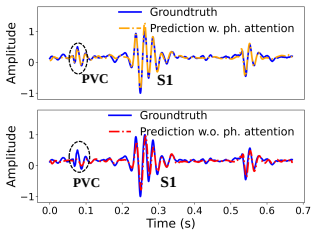


Figure 19: W.f. reconstructed w. and w.o. ph. attention.

faster convergence, achieving a much lower loss after the first epoch (2.76 vs. 19.4) and reaching its final loss value approximately 200 epochs earlier than the baseline. Furthermore, the PINN-based model shows superior generalization, achieving a final validation loss of 0.0795, a 26.5 % improvement over the baseline's 0.1083. The smaller gap between its training and validation losses (0.023 vs. 0.0418) also indicates reduced overfitting, a common benefit of physics-based regularization [59].

Crucially, these training advantages translate directly to improved reconstruction performance on unseen data. As shown in Figure 17, incorporating PINNs improved all evaluation metrics for a representative unseen subject. For instance, RMSE improved by 1.2 % and SSIM by 0.9 %. In summary, the ablation study confirms that PINNs not only accelerate model convergence but, more importantly, enhance their generalization ability, leading to more accurate and reliable PCG reconstruction.

5.3.2 Performance With and Without the Compensation Network
We next conduct an ablation study to evaluate the impact of the compensation network, following the same methodology as the experiments in Section 5.3.1. The results, depicted in Figure 16, demonstrate several key benefits. First, the compensation network significantly accelerates the convergence of the boundary loss and enables the model to reach a lower final loss value (from an average

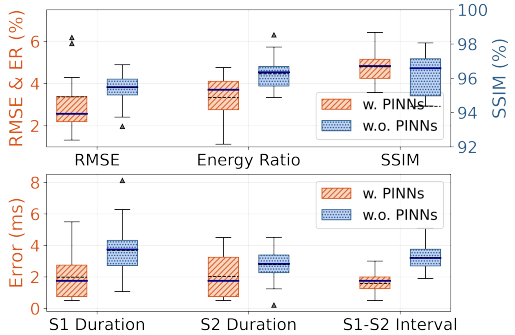


Figure 17: The error distribution of evaluation metrics w. and w.o. PINNs.

of 2.26 to 0.0047). Furthermore, it also reduces the PDE residual loss (from an average of 0.0048 to 0.0009), effectively balancing its magnitude with that of the boundary loss. Results successfully demonstrate the effectiveness design of the compensation network.

5.3.3 Performance With and Without the Fourier Feature Layer
Following the above experiments, we further explore the impact of FFL on the network performance. As shown in Figure 18, the model incorporating a FFL successfully captures the rapid, transient features of both the S1 and S2 components of PCG. In contrast, the baseline model without FFL suffers from significant waveform blunting. This is particularly evident in the S1 region, where its output waveform is considerably wider than the ground-truth signal from the apex, and it fails to reconstruct the S2 component entirely. Overall, the baseline model exhibits a characteristic over-smoothing of the signal. This comparison demonstrates that FFL effectively enhances the network's capacity to represent high-frequency information, thereby improving the inversion fidelity for the transient characteristics of PCG.

5.3.4 Performance With and Without Physical-guided Attention
We finally evaluate the impact of our proposed physical-guided attention loss, which is designed to force the model to focus on diagnostically critical regions of PCG. The results, shown in Figure 19, reveal the baseline's limitations. Without the attention mechanism, the model struggles to accurately reconstruct key pathological features, leading to significant amplitude attenuation in PVC events and timing errors in the S1 heart sound. In contrast, the physical-guided attention loss yields substantial improvements. By directing the model's focus, it reduces the local RMSE in the PVC region from 0.098 to 0.025 and decreases the S1 peak timing error from 2.0 ms to just 0.75 ms. This confirms that the physical-guided attention is crucial for enhancing model's ability to identify and precisely align key diagnostic features, thereby improving reconstruction fidelity.

5.4 Usability Study

To validate the usability of EarPCG in clinical practice, we design a subjective evaluation experiment. Experienced clinicians are asked to score the quality of reconstructed PCG based on ground truth

to assess their utility in medical diagnosis. To mitigate subjective bias, the audio samples are presented anonymously and in random order. After listening, clinicians rate the reconstructed PCG based on the following dimensions: 1) **Audio Similarity** – the overall perceptual resemblance between the reconstructed and ground truth; 2) **Preservation of Diagnostic Features** – whether key medical characteristics such as S1/S2 sounds, murmurs, and rhythm abnormalities are clearly identifiable and preserved; and 3) **Clinical Interpretability** – whether the reconstructed sound possess value for preliminary assessment during clinical auscultation. The score ranges from 0 to 5.

A total of 160 reconstructed PCG audio clips, each lasting for 20 s, are played on a Thinkpad X1 Carbon laptop. The employed doctors wear a BOSE SoundTrue Ultra earphone to listen to the PCG for evaluation. Based on expert evaluation, the average scores on the three dimensions of "Audio Similarity," "Preservation of Diagnostic Features," and "Clinical Interpretability" are 4.9 ± 0.25, 4.82 ± 0.14, and 4.89 ± 0.30, respectively. The high overall scores and small variance indicate that expert generally considers the reconstructed PCG to be highly consistent with those obtained by traditional stethoscopes. Results successfully demonstrate good clinical application prospects of EarPCG.

6 Related Work and Discussions

Earable sensing has recently emerged as a compelling platform for cardiac monitoring, attractive for its non-invasive nature and ubiquity compared to conventional systems [7, 27, 68]. Current approaches are generally categorized into active and passive sensing.

Active sensing methods primarily rely on emitting signals into the ear canal and analyzing the reflected signals modulated by physiological activities. For instance, the system proposed in APG [18] utilizes ultrasound emitted into the ear canal to detect volumetric changes modulated by vascular deformation, akin to PPG sensing, thereby indirectly estimating heart rate (HR) and HRV. This approach achieves estimation errors of 3.21 % and 2.70 % for HR and HRV, respectively, even during subject motion. Similar methods have also been proposed in Earmonitor [57].

Passive sensing, on the other hand, involves the earphone's built-in microphone listening for passive in-ear audio related to human physiological activities [6, 9, 23, 24, 34]. These systems are built on the premise that the occlusion effect [35] of a sealed ear canal has adequate passive gain to sense faint bone-conducted body sounds. Based on this principle, the authors of EarAce [9] posit that bone-conducted PCG can be detectable within an ear canal. They hence customize a versatile acoustic sensing platform based on commodity ANC earphones that is capable of extracting cardiac activity-related indicators (such as systole and diastole) under various wearing conditions and motion interference. They report median errors of 4.77% and 2.95% in systolic and diastolic period monitoring, respectively. This principle is also applied by the authors from HearBP [71] and hEARt [6] to estimate blood pressure and heart rate. hEARt achieves a resting heart rate monitoring accuracy of 3.02 ± 2.97 Beats Per Minute (BPM). And HearBP reports standard deviation errors of 3.13mmHg and 3.56mmHg for diastolic and systolic blood pressure measurements.

The work in [14] also claims to be able to extract PCG from in-ear canal. As bone-conducted PCG is subtle, due to frequency and energy distortions, they hence design customized hardware, including active amplifiers and impedance matching circuits, to improve sensitivity. In addition, a deep neural network is employed to compensate for those distortions. However, this body of passive sensing work rests on a questionable physical foundation. This more fundamental issue, which we identify in our own measurements, is a significant temporal latency (100 ms) between the in-ear signal and the true PCG. This latency is incompatible with the speed of sound through bone and challenges the core premise of all prior passive sensing work.

In contrast to prior work, EarPCG resolves this discrepancy by re-interpreting the in-ear signal not as a degraded PCG, but as a local measurement of the arterial pulse wave (an acoustic PPG). We note that although initial studies [20, 62] suggest that in-ear audio originates from vascular movement, they lack a comprehensive model of the underlying mechanism. In contrast, we develop a physics-informed model that accurately describes the signal's propagation and transduction, and we employ PINNs to solve the inverse problem of reconstructing true PCG. While our results demonstrate strong potential for continuous cardiac monitoring, we acknowledge limitations: our system has been validated only under static, moderate noise conditions. Future work will focus on improving robustness in dynamic environments, likely by incorporating multi-modal sensing with inertial measurement units and external microphones, to address motion artifacts and strong reverberations. Meanwhile, the current investigation is conducted with a limited number of participants, and thus, the generalizability of our findings may be constrained. To ascertain the clinical significance of our approach, extensive validation on a more diverse and larger patient population is required.

7 Conclusion

This paper presented EarPCG, a novel system for continuous cardiac monitoring using in-ear audio. We established a new physical model that interprets the in-ear signal as an acoustic PPG, resolving inconsistencies in prior work. By leveraging physics-informed neural network to invert this model, we successfully reconstructed high-fidelity PCG waveforms from commodity earphones. Our evaluation demonstrated that EarPCG achieves high reconstruction accuracy (RMSE: 2.935%, SSIM: 97.148%) across diverse subjects and accurately captures key clinical features, including the timing of cardiac events and signatures of pathology. These findings validate the potential of earable devices as a viable and powerful platform for clinical cardiac monitoring.

Acknowledgment

We are grateful to the anonymous shepherd and reviewers for their valuable comments. This research is supported by the Fundamental Research Funds for the Central Universities YCJJ20252406.

References

[1] 2022. 9 - Physical and physiological interpretations of the PPG signal. In *Photoplethysmography*, John Allen and Panicos Kyriacou (Eds.). Academic Press,

319–340.
[2] Rafi u Shan Ahmad, Muhammad Shehzad Khan, Mohamed Elhousseini Hilal, Bangul Khan, Yuanting Zhang, and Bee Luan Khoo. 2025. Advancements in wearable heart sounds devices for the monitoring of cardiovascular diseases. *SmartMat* 6, 1 (2025), e1311.
[3] Shovan Barma, Bo-Wei Chen, Wen Ji, Feng Jiang, and Jhing-Fa Wang. 2015. Measurement of duration, energy of instantaneous frequencies, and splits of subcomponents of the second heart sound. *IEEE Transactions on Instrumentation and Measurement* 64, 7 (2015), 1958–1967.
[4] Jesse Bettencourt, Matthew J Johnson, and David Duvenaud. 2019. Taylor-mode automatic differentiation for higher-order derivatives in JAX. In *Program Transformations for ML Workshop at NeurIPS 2019*.
[5] GARY W BURGGRAF and ERNEST CRAIGE. 1974. The first heart sound in complete heart block: phono-echocardiographic correlations. *Circulation* 50, 1 (1974), 17–24.
[6] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2023. heart: Motion-resilient heart rate monitoring with in-ear microphones. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 200–209.
[7] Yetong Cao, Chao Cai, Fan Li, Zhe Chen, and Jun Luo. 2023. HeartPrint: Passive Heart Sounds Authentication Exploiting In-Ear Microphones. *Heart* 50, S1 (2023), S2.
[8] Yetong Cao, Chao Cai, Fan Li, Zhe Chen, and Jun Luo. 2024. Enabling Passive User Authentication via Heart Sounds on In-Ear Microphones. *IEEE Transactions on Dependable and Secure Computing* (2024).
[9] Yetong Cao, Chao Cai, Anbo Yu, Fan Li, and Jun Luo. 2023. EarAce: Empowering Versatile Acoustic Sensing via Earable Active Noise Cancellation Platform. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–23.
[10] Kévin Carillo, Olivier Doutres, and Franck Sgard. 2020. Theoretical investigation of the low frequency fundamental mechanism of the objective occlusion effect induced by bone-conducted stimulation. *The Journal of the Acoustical Society of America* 147, 5 (2020), 3476–3489.
[11] Yong-Mei Cha, Glenn K Lee, Kyle W Klarich, and Martha Grogan. 2012. Premature ventricular contraction-induced cardiomyopathy: a treatable condition. *Circulation: Arrhythmia and Electrophysiology* 5, 1 (2012), 229–236.
[12] chaocai. 2023. EarAce project. https://github.com/caichao/earace Last accessed: 2025-4-1.
[13] Tuochao Chen, Malek Itani, Sefik Emre Eskimez, Takuya Yoshioka, and Shyamnath Gollakota. 2024. Hearable devices with sound bubbles. *Nature Electronics* (2024), 1–12.
[14] Tao Chen, Yongjie Yang, Xiaoran Fan, Xiuzhen Guo, Jie Xiong, and Longfei Shangguan. 2024. Exploring the Feasibility of Remote Cardiac Auscultation Using Earphones. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 357–372.
[15] Zhihua Chen, An Huang, and Xiaoli Qiang. 2020. Improved neural networks based on genetic algorithm for pulse recognition. *Computational Biology and Chemistry* 88 (2020), 107315.
[16] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. 2021. MoVi-Fi: Motion-robust vital signs waveform recovery via deep interpreted RF sensing. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 392–405.
[17] Maurice Enriquez-Sarano, Cary W Akins, and Alec Vahanian. 2009. Mitral regurgitation. *The Lancet* 373, 9672 (2009), 1382–1394.
[18] Xiaoran Fan, David Pearl, Richard Howard, Longfei Shangguan, and Trausti Thormundsson. 2023. APG: Audioplethysmography for Cardiac Monitoring in Hearables. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–15.
[19] Luca Formaggia, Daniele Lamponi, and Alfio Quarteroni. 2003. One-dimensional models for blood flow in arteries. *Journal of engineering mathematics* 47 (2003), 251–276.
[20] Bjarke Gaardbaek and Preben Kidmose. 2024. On the origin of cardiovascular sounds recorded from the ear. *IEEE Transactions on Biomedical Engineering* 72, 1 (2024), 210–216.
[21] Francis Roosevelt Gilliam III, Robert Ciesielski, Karlen Shahinyan, Pratistha Shakya, John Cunsolo, Jal Mahendra Panchal, Bartlomiej Król-Józaga, Monika Król, Olivia Kierul, Charles Bridges, et al. 2022. In-ear infrasonic hemodynography with a digital health device for cardiovascular monitoring using the human audiome. *NPJ Digital Medicine* 5, 1 (2022), 189.
[22] Unsoo Ha, Salah Assana, and Fadel Adib. 2020. Contactless seismocardiography via deep learning radars. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[23] Feiyu Han, Panlong Yang, Yuanhao Feng, Weiwei Jiang, Youwei Zhang, and Xiang-Yang Li. 2024. Earsleep: In-ear acoustic-based physical and physiological activity recognition for sleep stage detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–31.
[24] Changshuo Hu, Thivya Kandappu, Yang Liu, Cecilia Mascolo, and Dong Ma. 2024. BreathPro: Monitoring breathing mode during running with earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–25.
[25] Changshuo Hu, Xiao Ma, Xinger Huang, Yiran Shen, and Dong Ma. 2024. LR-Auth: Towards Practical Implementation of Implicit User Authentication on Earbuds. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 4 (2024), 1–27.
[26] Quincy A Huhn, Mauricio E Tano, and Jean C Ragusa. 2023. Physics-informed neural network with fourier features for radiation transport in heterogeneous media. *Nuclear Science and Engineering* 197, 9 (2023), 2484–2497.
[27] Yincheng Jin, Yang Gao, Xiaotao Guo, Jun Wen, Zhengxiong Li, and Zhanpeng Jin. 2022. Earhealth: an earphone-based acoustic otoscope for detection of multiple ear diseases in daily life. In *Proceedings of the 20th annual international conference on mobile systems, applications and services*. 397–408.
[28] IH Jung, SA Bae, MK Kim, IK Moon, HS Seo, and HJ Chang. 2024. Machine learning approach for detection of valvular heart disease through digital cardiac auscultation: results from multi-center prospective cross-sectional study. *European Heart Journal* 45, Supplement_1 (2024), ehae666–3447.
[29] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. 2021. Physics-informed machine learning. *Nature Reviews Physics* 3, 6 (2021), 422–440.
[30] Hans Petter Langtangen. 2016. Finite difference methods for wave motion. *Center for Biomedical Computing, Simula Research Laboratory 2Department of Informatics, University of Oslo* (2016), 5–9.
[31] Sung Hoon Lee, Yun Soung Kim, and Woon-Hong Yeo. 2023. Fully Portable Wireless Soft Stethoscope and Machine Learning for Continuous Real-Time Auscultation and Automated Disease Detection. In *2023 IEEE 73rd Electronic Components and Technology Conference (ECTC)*. IEEE, 1433–1437.
[32] Yongjoon Lee and Chanwoo Kim. 2024. Wave-u-mamba: an end-to-end framework for high-quality and efficient speech super resolution. *arXiv preprint arXiv:2409.09337* (2024).
[33] Shengping Liu, Guanlan Chen, and Guoming Chen. 2013. The Application of Wavelet Analysis and BP Neural Network for the Early Diagnosis of Coronary Heart Disease. In *Emerging Technologies for Information Systems, Computing, and Management*. Springer, 165–172.
[34] Yang Liu, Kayla-Jade Butkow, Jake Stuchbury-Wass, Adam Pullin, Dong Ma, and Cecilia Mascolo. 2024. RespEar: Earable-Based Robust Respiratory Rate Monitoring. *arXiv preprint arXiv:2407.06901* (2024).
[35] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. Oesense: employing occlusion effect for in-ear human sensing. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 175–187.
[36] Flavius Gabriel Marc. 2023. *Feasibility of monitoring congestive heart failure with seismocardiography: a literature review*. Master's thesis. Utrecht University. Available from Utrecht University repository.
[37] Rainer Martin. 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on speech and audio processing* 9, 5 (2001), 504–512.
[38] Sheng MIAO and Zhong Lihui. Jian'e Dong, Jingyu Hou. 2020. Spectrum Characteristics Analysis and Recognition of CHD Heart Sound in Five Auscultation Locations. *Journal of Biomedical Engineering and Technology*. 8, 1 (2020), 14–24.
[39] World Health Organization. 2025. Cardiovascular diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed: 2025-01-06.
[40] Ebru Ozbilge. 2013. Determination of the unknown boundary condition of the inverse parabolic problems via semigroup method. *Boundary Value Problems* 2013 (2013), 1–7.
[41] Guofei Pang, Lu Lu, and George Em Karniadakis. 2019. fPINNs: Fractional physics-informed neural networks. *SIAM Journal on Scientific Computing* 41, 4 (2019), A2603–A2626.
[42] Ronald L Panton. 2024. *Incompressible flow*. John Wiley & Sons.
[43] H Parsaei, A Vakily, and AM Shafiei. 2017. A wireless electronic esophageal stethoscope for continuous monitoring of cardiovascular and respiratory systems during anaesthesia. *Journal of Biomedical Physics & Engineering* 7, 1 (2017), 69.
[44] Cristhian Potes, Saman Parvaneh, Asif Rahman, and Bryan Conroy. 2016. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In *2016 computing in cardiology conference (CinC)*. 621–624.
[45] Jay Prakash, Zhijian Yang, Yu-Lin Wei, Haitham Hassanieh, and Romit Roy Choudhury. 2020. EarSense: earphones as a teeth activity sensor. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.
[46] Stein Inge Rabben, Nikos Stergiopulos, Leif Rune Hellevik, Otto A Smiseth, Stig Slørdahl, Stig Urheim, and Bjørn Angelsen. 2004. An ultrasound-based method for determining pulse wave velocity in superficial arteries. *Journal of biomechanics* 37, 10 (2004), 1615–1622.
[47] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* 378 (2019), 686–707.

[48] Sridhar Ramakrishnan, Satish Udpa, and Lalita Udpa. 2009. A numerical model simulating sound propagation in human thorax. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 530–533.
[49] Kay Raum, Ingrid Leguerney, Florent Chandelier, Emmanuel Bossy, Maryline Talmant, Amena Saïed, Françoise Peyrin, and Pascal Laugier. 2005. Bone microstructure and elastic tissue properties are reflected in QUS axial transmission measurements. *Ultrasound in medicine & biology* 31, 9 (2005), 1225–1235.
[50] Todd R Reed, Nancy E Reed, and Peter Fritzson. 2004. Heart sound analysis for symptom detection and computer-aided diagnosis. *Simulation Modelling Practice and Theory* 12, 2 (2004), 129–146.
[51] Francesco Renna, Jorge Oliveira, and Miguel T Coimbra. 2019. Deep convolutional neural networks for heart sound segmentation. *IEEE journal of biomedical and health informatics* 23, 6 (2019), 2435–2445.
[52] Lillian M Rigoli, Daniel Holman, Michael J Spivey, and Christopher T Kello. 2014. Spectral convergence in tapping and physiological fluctuations: coupling and independence of 1/f noise in the central and autonomic nervous systems. *Frontiers in human neuroscience* 8 (2014), 713.
[53] Jason H Rogers and Steven F Bolling. 2009. The tricuspid valve: current perspective and evolving management of tricuspid regurgitation. *Circulation* 119, 20 (2009), 2718–2725.
[54] Fred Shaffer and Jay P Ginsberg. 2017. An overview of heart rate variability metrics and norms. *Frontiers in public health* 5 (2017), 258.
[55] Yalda Shahriari, Richard Fidler, Michele M Pelter, Yong Bai, Andrea Villaroman, and Xiao Hu. 2017. Electrocardiogram signal quality assessment based on structural image similarity metric. *IEEE Transactions on Biomedical Engineering* 65, 4 (2017), 745–753.
[56] Stefan Stenfelt and Sabine Reinfeldt. 2007. A model of the occlusion effect with bone-conducted stimulation. *International journal of audiology* 46, 10 (2007), 595–608.
[57] Xue Sun, Jie Xiong, Chao Feng, Wenwen Deng, Xudong Wei, Dingyi Fang, and Xiaojiang Chen. 2023. Earmonitor: In-ear motion-resilient acoustic sensing using commodity earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–22.
[58] W Reid Thompson. 2017. In defence of auscultation: a glorious future? *Heart Asia* 9, 1 (2017), 44–47.
[59] Tom Viering and Marco Loog. 2022. The shape of learning curves: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 6 (2022), 7799–7819.
[60] Jiaming Wang, Tao You, Kang Yi, Yaqin Gong, Qilian Xie, Fei Qu, Bangzhou Wang, and Zhaoming He. 2020. Intelligent diagnosis of heart murmurs in children with congenital heart disease. *Journal of healthcare engineering* 2020, 1 (2020), 9640821.
[61] Sifan Wang, Xinling Yu, and Paris Perdikaris. 2022. When and why PINNs fail to train: A neural tangent kernel perspective. *J. Comput. Phys.* 449 (2022), 110768.
[62] Jordan Waters, Jake Stuchbury-Wass, Yang Liu, Kayla-Jade Butkow, and Cecilia Mascolo. 2024. Deep-learning based segmentation of in-ear cardiac sounds. (2024).
[63] YU WELL. 2023. The details about YUWELL versatile stethoscope. https://yuyueshop.com/goods/2618781.html Accessed: 2025-06-9.
[64] GL Woolam, PL Schnur, C Vallbona, and HE Hoff. 1962. The pulse wave velocity as an early indicator of atherosclerosis in diabetic subjects. *Circulation* 25, 3 (1962), 533–539.
[65] Chenhan Xu, Tianyu Chen, Huining Li, Alexander Gherardi, Michelle Weng, Zhengxiong Li, and Wenyao Xu. 2022. Hearing Heartbeat from Voice: Towards Next Generation Voice-User Interfaces with Cardiac Sensing Functions. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, SenSys 2022, Boston, Massachusetts, November 6-9, 2022*. 149–163.
[66] Chenhan Xu, Huining Li, Zhengxiong Li, Hanbin Zhang, Aditya Singh Rathore, Xingyu Chen, Kun Wang, Ming-chun Huang, and Wenyao Xu. 2021. Cardiacwave: A mmwave-based scheme of non-contact and high-definition heart activity computing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–26.
[67] Mustafa Yamacli, Zumray Dokur, and Tamer Olmez. 2008. Segmentation of S1–S2 sounds in phonocardiogram records using wavelet energies. In *2008 23rd International Symposium on Computer and Information Sciences*. IEEE, 1–6.
[68] Zhijian Yang and Romit Roy Choudhury. 2021. Personalizing head related transfer functions for earables. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*. 137–150.
[69] John M Zanetti and Kouhyar Tavakolian. 2013. Seismocardiography: Past, present and future. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. 7004–7007.
[70] Shujie Zhang, Tianyue Zheng, Zhe Chen, and Jun Luo. 2022. Can we obtain fine-grained heartbeat waveform via contact-free RF-sensing?. In *IEEE INFOCOM 2022-IEEE conference on computer communications*. IEEE, 1759–1768.
[71] Zhiyuan Zhao, Fan Li, Yadong Xie, Huanran Xie, Kerui Zhang, Li Zhang, and Yu Wang. 2024. HearBP: Hear Your Blood Pressure via In-ear Acoustic Sensing Based on Heart Sounds. In *IEEE INFOCOM 2024-IEEE Conference on Computer Communications*. IEEE, 991–1000.
[72] Zongren Zou, Xuhui Meng, and George Em Karniadakis. 2024. Correcting model misspecification in physics-informed neural networks (PINNs). *J. Comput. Phys.* 505 (2024), 112918.