ECG Signal Construction From Heart Sounds via Single Node, Surface Acoustic Sensing

Kaylee Yaxuan Li, Yasha Iravantchi, Hyunmin Park, Yiming Liu, Alanson Sample
Computer Science & Engineering, Univerisity of Michigan, Ann Arbor, USA
{yaxuanli, yiravan, phyunmin, liuym, apsample}@umich.edu

***Abstract*—Effective means of enabling single-lead, non-intrusive, and dry electrocardiogram (ECG) measurements offer the potential for prolonged cardiac rhythm monitoring of mobile users in non-clinical environments. However, existing ECG measurement approaches require accurate electrode placement, cumbersome wiring, and require users to be stationary. Alternatively, current heart sound-based approaches such as phonocardiograms lack the sensitivity and precision to detect crucial cardiac rhythm features and are vulnerable to environmental noise. This work utilizes a wide bandwidth surface-acoustic-wave microphone on the neck to capture heart sounds via the carotid artery. A cross-modal autoencoder, a state-of-the-art algorithm for signal modality conversion, is proposed to transform heart acoustic signals into corresponding ECG waveforms. Results from a 9 participant study demonstrate the effectiveness of constructing a PQRST waveform from acoustic heart sounds and accurately determining critical PQRST metrics. Finally, mobile acoustic ECG wave construction of a user walking is demonstrated, laying the groundwork for unobtrusive, long-term, low-cost daily cardiac rhythm monitoring.**

***Clinical relevance*—Transforming heart sound signals to produce prominent ECG metrics enables low-cost daily cardiac rhythm monitoring using a single-node dry wearable device.**

I. Introduction

Per the World Health Organization (WHO), approximately 17.9 million deaths annually are due to cardiovascular diseases [1]. The Electrocardiogram (ECG), which measures the electric potential of the heart's atrial and ventricular activities, is the most commonly used technique of representing cardiac rhythm for diagnosing abnormal situations such as arrhythmia and heart morphology [2]. The ECG signal is characterized by the PQRST complex wave. The P-wave represents atrial depolarization, the QRS complex depicts ventricular depolarization, and the T-wave indicates ventricular repolarization. Analyzing PQRST complex enables physicians to make precise and comprehensive cardiovascular diagnoses.

However, conventional ECG devices necessitate precise placement of multiple nodes on the body [3], [4], usually requiring the assistance of professional expertise within a clinical setting. Consequently, the development of accessible, cost-effective, and easy to use ECG detection methods is crucial. Alternatively, stethoscopes have been utilized to collect heart sounds or phonocardiograms (PCG), providing a more generalized depiction of cardiac rhythm than ECG. Nevertheless, stethoscopes continue to face challenges from environmental noise interference and respiratory sounds.

In response to these challenges, this research proposes to employ a 48kHz sampling-rate surface acoustic wave microphone ($2 USD), positioned on the neck, to capture the carotid artery's sound induced by heart movements as shown in Figure 1. Leveraging a state-of-the-art cross-modal autoencoder machine learning model, the heart acoustic signals are transformed into ECG signals. The evaluation focuses on seven crucial PQRST features. This research represents an initial step towards demonstrating the feasibility of converting PCG signals into ECG signals. It enables the development of a dry, single-node, low-cost device for long-term monitoring of heart rhythm in everyday settings.

II. Related Work

ECG signal measurements: Electrocardiogram (ECG) technology, employed for decades in medical and consumer applications, aids in diagnosing diseases and monitoring heart function [5]. Clinical ECG typically utilizes 8-12 leads for high-fidelity and low-noise measurements [6]. It requires placing electrical leads with conductive gels at specific body locations, including on the chest, arms, and legs [1], [6]. This process usually requires a trained technician, with the patient lying down and still for data collection. Clinical ECG devices, limited in home use, have motivated development towards mobile, accessible alternatives for general users.

Consumer-grade ECG devices, which often use dry leads, require noise filtration, resampling, and data normalization due to factors such as baseline shifts, muscular and instrumental noise. Feature extraction algorithms are employed, followed by the application of machine learning models to identify abnormal cardiac rhythms. However, these devices may compromise data precision and fidelity vs. deployability [7].

Cross-modality learning with Autoencoders: Recently, researchers have focused on converting information across different modalities, such as generating images from text [8]. Initially designed for input data reconstruction [9], autoencoders have shown great performance in various applications such as feature extraction, abnormality detection, and biological noise reduction. More recently, autoencoders have evolved from extracting salient features of one data modality [10] to facilitating knowledge adaptation and transfer from one modality to another. Utilizing advanced machine learning techniques, researchers have employed cross-modal autoencoders to produce cardiac MRIs from ECGs [11] and have used

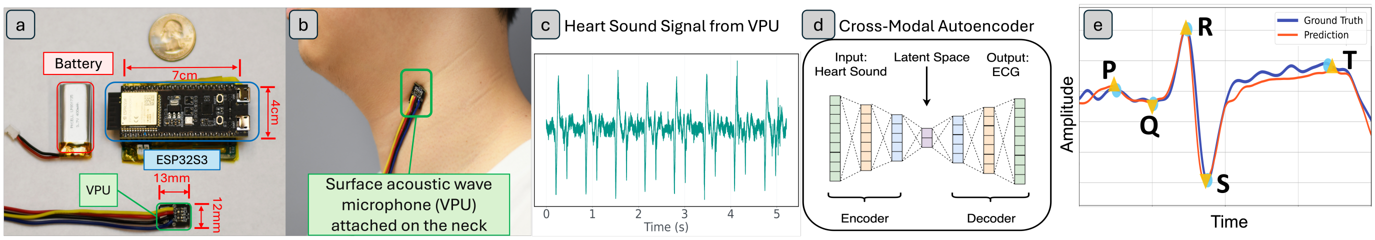

Fig. 1. A conceptual overview of the Surface Acoustic Wave (SAW) to ECG signal construction pipeline consisting of: (a) A custom designed PCB with an ESP32 microcontroller which is wired to the SAW microphone and powered by a 3.7V battery. (b) A user with the SAW microphone placed on their neck. (c) Captured raw heart sound signal. (d) the ESP32S3 microcontroller transmits data wirelessly to the computer, and a cross-modal autoencoder is used for ECG signal construction. (e) Resulting ECG ground truth and construction signals, along with PQRST data points indicated.

generative adversarial networks (GANs) to generate multi-lead ECGs from single-lead ECG input data [12]. However, cross-modality learning in biological signals is still in the early stages, with limited research compared to traditional image and text modalities [13]. Hence, this work, employing a cross-modal autoencoder for modality conversion, establishes the foundations in demonstrating the feasibility of transforming physiological waves from enhanced heart sound signals into ECG signals.

III. Materials and Methods

This section describes hardware design, the data collection and preprocessing pipeline, and the architecture of the cross-modal autoencoder machine learning model.

A. Hardware Design

The surface acoustic wave microphone, also known as a Sonion Voice PickUp (VPU) sensor, is developed to isolate the user's voice via body transmission. It can pick up sounds from the contacted body surface while rejecting ambient noise from the environment, as shown in previous works [14]. These devices employ a standard Pulse Density Modulation (PDM) protocol for digital data transmission and multiple channels can be synchronized with a shared clock. As displayed in Figure 1(a), ESP32-S3 microcontroller was chosen for processing, noted for its integrated hardware peripheral [15] that can efficiently decode multi-channel PDM efficiently. A 3.7V 400mAh lithium-ion battery supports approximately 5 hours of continuous operation, allowing for fully wireless functionality. The device steams 16-bit 48kHz audio over WiFi to a laptop computer for data collection.

B. Ground Truth Data Acquisition

An open-source ECG platform (SHIELD-EKG-EMG [16]) was employed to capture the differential bio-potentials of the heart and digitize the signal, providing the ECG ground truth data. A USB Serial connection transmits the ECG data to a laptop computer as a 10-bit 500Hz stream. The heart's acoustic signal captured via VPU, and the ECG ground truth are time-synchronized using the FFmpeg library [17] for comparative analysis.

C. Dataset Collection and Data Pre-processing

To evaluate the machine learning model for reconstruction, 9 participants (4 female and 5 male, mean age = 23.9, SD = 2.70) collected ECG and heart sound datasets, in accordance with our Institutional Review Board. The ECG Red Dot Cloth electrodes [18] were pasted on each participant's left and right wrist, and left ankle for ground truth data collection using the aforementioned development kit per the directions of the device. Concurrently, the participant held the VPU against their neck to capture the heart's acoustic signal synchronously, as seen in Figure 1(b). The neck carotid artery was selected for microphone placement due to its common acceptance, user comfort, and ease of data collection compared to other heartbeat-detectable positions. It should be noted, that the microphone was also able to collect heart sounds successfully in typical PCG placements, such as on the apex cordis. Participants were instructed to lie on a chair and remain still during the data collection, per typical ECG/PCG data collection procedures. A total of 20 minutes of data were collected from each user. The heart acoustic signal example is presented in Figure 1(c).

Upon acquiring the data, it was segmented into two sets of windows: one capturing every two consecutive heartbeats and another for individual heartbeats. Each heartbeat consists of one systolic contraction and one diastolic relaxation of the heart. Since all participants were healthy individuals without any reported heart-related diseases, each heartbeat is expected to encompass one PQRST complex. R-dot was annotated using NeuroKit2 library [19], and then aligned the data samples either one R or an RR interval at the center of each sample. Subsequently, normalization was applied to a 0-1 range. In the end, 12,150 times single heartbeat data samples were collected from 9 participants with synchronized ECG and PCG signals. For later individualized model performance evaluation, each user's data was split, allocating 80% for training purposes and reserving 20% as the testing set. This dataset will be publicly available following the completion of all necessary permission procedures and supervisory protocols.

D. Cross-modal Autoencoder Model

The ESP32-S3 microcontroller acquires data and transmits it to a laptop via WiFi. Concurrently, a machine learning

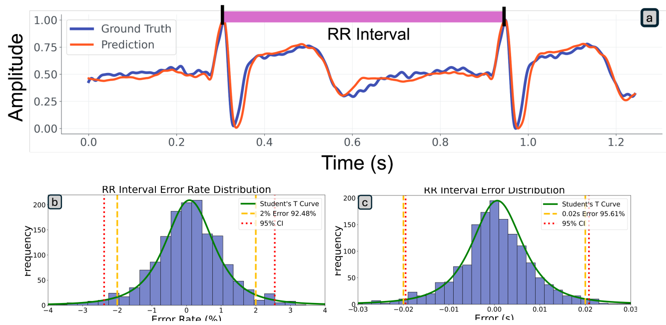

Fig. 2. (a): Constructed and ground truth ECG signal. RR interval relative (b) and absolute error (c) frequency distribution

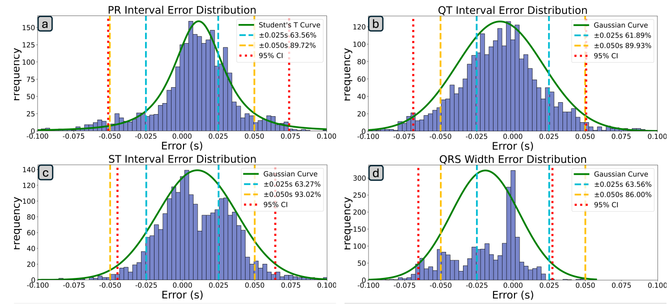

Fig. 3. Frequency distribution of error in the PR (a), ST (b), QT (c) intervals, and QRS width (d) with corresponding fitting distribution models. The bin width is 0.003 seconds. Error boundaries at ±0.025 seconds and ±0.050 seconds, along with the 95% confidence interval (CI), are indicated.

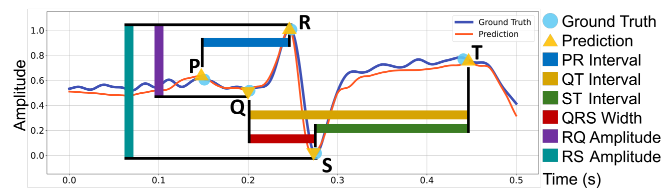

Fig. 4. Overlay of ECG signal construction/prediction with ground truth across time, showing PR, QT, ST intervals, QRS width, and RQ, RS amplitude.

model (cross-modal autoencoder) is employed to construct corresponding ECG signals from heart sounds captured by the VPU. Here presents the architecture of the machine learning model as illustrated in Figure 1(d). The cross-modal autoencoder is composed of an encoder and a decoder. The encoder's function is to compress the input, extract key features, and learn all important information about the data. Subsequently, the decoder attempts to create the corresponding modality data from salient features. The encoder utilizes a stack of 1D convolutional layers with LeakyReLU activation, each followed by MaxPooling for dimensionality reduction and Dropout layers for regularization to prevent model overfitting, creating a compressed representation of the input data in latent space. Correspondingly, the decoder mirrors this architecture, employing convolutional layers and upsampling layer to create the corresponding ECG signal. The model is compiled using the Adam optimizer and mean squared error loss function, ensuring efficient learning and error minimization.

IV. Results and Discussion

The time intervals and amplitude ranges based on an ECG's PQRST complex have diagnostic implications; variations from the typical morphology may suggest pathological conditions. Consequently, this section presents formulated evaluation metrics based on clinically significant ECG characteristics [6].

This research aims to demonstrate the possibility and establish the groundwork for using a user-dependent cross-modal autoencoder to transform heart acoustic signals into ECG features, as shown in Figure 1(e), and presents seven critical evaluation metrics adopted to assess the constructed ECG (i.e., from heart sounds) against the ground truth ECG.

A. RR interval metric

The RR interval, delineating the duration between successive R peaks, is a prevalent diagnostic metric reflective of cardiac rhythm. Deviations in the regularity of RR interval can indicate pathophysiological conditions such as atrioventricular block or atrial fibrillation [6]. A cross-modal autoencoder model was employed to reconstruct the ECG signal from an acoustic signal. The evaluation compared RR intervals from the autoencoder-constructed ECG and the ground truth, as shown in Figure 2(a). The mean absolute error and mean relative error (defined as $((pred - gt)/gt)$ across all users are calculated, also depicted in Figure 2(b,c). The mean relative error typically ranged within ±4%. The mean absolute error exhibits 95.61% of predictions have at most ±0.02s error compared to ground truth. The Student's T distribution is utilized to model the error frequency distribution, highlighting that a considerable portion of the data approaches zero error.

B. PR, ST, QT interval and QRS width metrics

Considering the clinical importance of ECG signals, our study also emphasizes additional key metrics: the PR interval, ST interval, QT interval, and QRS width. For example, abnormalities in the PR interval can suggest atrioventricular block, complete atrioventricular block, or electrolyte imbalances, particularly hypokalemia [6]. As illustrated in Figure 4, the ground truth and the constructed signal closely align, indicating the cross-modal autoencoder model's strong performance in modality transform. Furthermore, the PQRST complexes in both signals, detected using the same analytical model [19], align well, further demonstrating the model's high efficacy. For a more comprehensive evaluation, mean absolute error is employed in the time domain to quantify performance across all users, as depicted in Figure 3. All four metrics have ±0.1s error at most. Taking PR interval as an example, 89.72% of collected instances exhibit an error margin of only 0.05s. Using the Student's T distribution, the distribution reveals a significant proportion of the data is clustered around zero error.

C. QRS complex amplitude difference metrics

RQ, RS amplitude difference is defined as R minus Q and S value. The amplitude of QRS offers insights into ventricular myocardial mass and electrical conduction. Elevated RQ

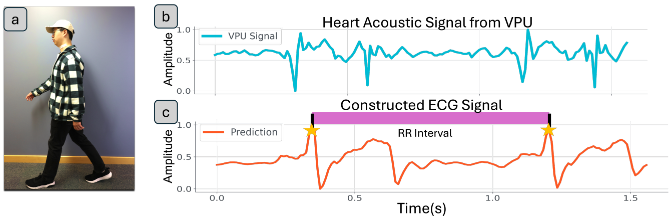

Fig. 5. (a). User wearing VPU on the neck is walking. (b). Raw heart sound collected from VPU. (c). Constructed ECG signal with RR interval indicated.

amplitudes may indicate left ventricular hypertrophy, whereas reduced values often point to right ventricular hypertrophy. The RS amplitude difference similarly plays a pivotal role in diagnosing bundle branch blocks [6]. In line with the preprocessing approach, where ground truth ECG data is normalized between 0 and 1, the focus is not on the true voltage values. Rather, this project aims to compare the model's accuracy in predicting RQ and RS differences, as shown in Figure 4. By comparing the predicted RQ and RS amplitude difference with the ground truth, prediction's accuracy is quantified as a percentage match to the actual values. The results demonstrate an 83% and 92% match in RQ and RS amplitude differences across all users, suggesting the model's ability to construct PQRST complex features.

D. Discussion and Limitations

Monitoring heart health in daily life, where movement involving, often conflicts with the continuous stillness and posture required by some ECG systems. To address this, preliminary experiments were conducted with participants using our system while walking shown in Figure 5. Recognizing that the ground truth ECG system demands user stillness for data collection, the pre-trained model of a single participant from formal user study is used, and applied it to acoustic heart sound data collected during a different session from that participant while walking. This data was used as test data for evaluation. As illustrated in Figure 5, raw acoustic signals from the VPU presents relatively more noise, yet the autoencoder manages to construct a reasonably accurate ECG signal. While initial results are promising, more rigorous experiments are needed to enhance robustness against real-world artifacts. With advancements in form-factor, better attachment methods, and larger datasets, our system could effectively monitor heart health during user movement.

This work transforming heart sounds into ECG signals primarily aims to establish a complementary diagnostic sensing method, rather than replace conventional ECG machines. Additionally, the current model is relatively modest in scale. Future developments in advanced machine learning models and larger datasets could enhance accuracy and enable users to employ the model directly, eliminating the need to supply their training data. Future work will investigate estimating the PQRST metrics locally on the device to reduce excessive data collection and enhance user privacy protection.

V. Conclusion

This study employs a 2.65 mm × 3.50 mm, wide bandwidth, highly sensitive surface-acoustic-wave-based microphone placed on the neck to collect heart sounds. Utilizing a cross-modal autoencoder model, we used these acoustic signals to construct ECG signals and find several ECG-related features can be accurately measured through a study with 9 participants. These findings demonstrate the feasibility of using a single-node, dry microphone and cross-modal autoencoder for monitoring ECG-related features in daily life.

References

[1] W. H. O. W. H. Ranking, "Cardiovascular diseases," 2020.
[2] T. Debnath, M. M. Hasan, and T. Biswas, "Analysis of ecg signal and classification of heart abnormalities using artificial neural network," in *2016 9th International Conference on Electrical and Computer Engineering (ICECE)*. IEEE, 2016, pp. 353–356.
[3] U. Satija, B. Ramkumar, and M. S. Manikandan, "A review of signal processing techniques for electrocardiogram signal quality assessment," *IEEE reviews in biomedical engineering*, vol. 11, pp. 36–52, 2018.
[4] B. Young, "New standards for ecg equipment," *Journal of electrocardiology*, vol. 57, pp. S1–S4, 2019.
[5] V. A. Ardeti, V. R. Kolluru, G. T. Varghese, and R. K. Patjoshi, "An overview on state-of-the-art electrocardiogram signal processing methods: Traditional to ai-based approaches," *Expert Systems with Applications*, p. 119561, 2023.
[6] B. Health, "Ecg study guide," 2016.
[7] N. Matsuura, K. Kuwabara, and T. Ogasawara, "Lightweight heartbeat detection algorithm for consumer grade wearable ecg measurement devices and its implementation," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 4299–4302.
[8] Z. Huang, X. Jin, C. Lu, Q. Hou, M.-M. Cheng, D. Fu, X. Shen, and J. Feng, "Contrastive masked autoencoders are stronger vision learners," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
[9] D. H. Ballard, "Modular learning in neural networks," in *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, ser. AAAI'87. AAAI Press, 1987, pp. 279–284. [Online]. Available: http://dl.acm.org/citation.cfm?id=1863696.1863746
[10] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa, "Deep learning in human activity recognition with wearable sensors: A review on advances," *Sensors*, vol. 22, no. 4, p. 1476, 2022.
[11] A. Radhakrishnan, S. F. Friedman, S. Khurshid, K. Ng, P. Batra, S. A. Lubitz, A. A. Philippakis, and C. Uhler, "Cross-modal autoencoder framework learns holistic representations of cardiovascular state," *Nature Communications*, vol. 14, no. 1, p. 2436, 2023.
[12] H.-C. Seo, G.-W. Yoon, S. Joo, and G.-B. Nam, "Multiple electrocardiogram generator with single-lead electrocardiogram," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106858, 2022.
[13] K.-K. Tseng, C. Wang, Y.-F. Huang, G.-R. Chen, K.-L. Yung, and W.-H. Ip, "Cross-domain transfer learning for pcg diagnosis algorithm," *Biosensors*, vol. 11, no. 4, p. 127, 2021.
[14] Y. Iravantchi, Y. Zhao, K. Kin, and A. P. Sample, "Sawsense: Using surface acoustic waves for surface-bound event recognition," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3544548.3580991
[15] ESP32, "Inter-ic sound (i2s) - esp32 - — esp-idf programming guide latest documentation," 2023.
[16] O. Ltd, "Shield-ekg-emg - open source hardware board," 2024.
[17] ffmpeg, "Ffmpeg," https://ffmpeg.org/.
[18] 3M, "3m red dot soft cloth monitoring electrode — 3m united states," https://www.3m.com/3M/en_US/p/d/b00037611/.
[19] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. A. Chen, "Neurokit2: A python toolbox for neurophysiological signal processing," *Behavior research methods*, pp. 1–8, 2021.