Heart Rate Monitoring Through ANC Headphones in Unconstrained Environments

[1]Zhenyu Wu, [1]Maanya Shanker, [2]Tao Chen, [3]Xiaoran Fan, [1,3]Longfei Shangguan
[1]*University of Pittsburgh,* [2]*Samsung Research America,* [3]*Google*

***Abstract*—This paper introduces CLEAR-APG, a novel acoustic sensing approach that enables reliable heart rate monitoring in unconstrained environments using off-the-shelf active noise cancellation (ANC) headphones. By emitting ultrasonic signals into the user's ear canal via the headphone speaker and analyzing their echoes, which can detect the frequency of a pulsating vein along the canal wall. However, everyday activities such as exercising, speaking, or eating cause jaw movements that deform the ear canal, overwhelming the subtle deformation caused by blood flowing. To overcome this challenge, we employ the ANC headphone's built-in gyroscope to capture body motion and identify how various motion patterns influence the heartbeat waveform. Building on this insight, we propose a multi-modal method that effectively denoises the heartbeat waveform measurements and further accurately extracts heart rate. We implement CLEAR-APG on ANC earbuds and conduct comprehensive field studies on 14 users. The results show that CLEAR-APG achieves an average heart rate error of 4.01% across seven different activities, satisfying industry-required margin of 10% heart rate error.**

I. Introduction

Continuous heart rate monitoring through wearable devices allows individuals to make informed health decisions and self-monitor their well-being. Medical professionals can analyze historical data spanning days, weeks, or even months to detect cardiovascular issues [1]. Additionally, long-term monitoring enables early detection of deteriorating trends in vital signs, allowing timely intervention [2].

The current state-of-the-art in wearable heart rate monitoring is the smartwatch, which uses a photoplethysmography (PPG) sensor [3]. The sensor emits red or green light into the skin, and the reflected signal varies with blood volume changes in the wrist's veins [4]. However, despite its convenience, PPG's accuracy is affected by skin tone, as darker skin with higher melanin content absorbs more green light, leading manufacturers to recommend using the device only at rest or on lighter skin tones [5], [6].

To address this issue, researchers have explored ear-based, acoustic heart rate monitoring solution. A more recent alternative, audioplethysmography (APG), has gained traction [7]. APG works by emitting an inaudible acoustic signal (i.e., ultrasound) into the ear canal and analyzing the reflections, which capture pulsations from a vein in the canal wall. Unlike PPG sensor, APG is not influenced by melanin levels. Additionally, since APG only requires a speaker and microphone pair, mobile users can readily implement it on their devices using standard active noise cancellation (ANC) headphones.

However, physical activities such as jogging, eating, talking, and exercising can cause significant deformation of the ear canal. These body-induced changes are orders of magnitude stronger than the subtle pulsations of the vein in the ear canal wall. Since APG relies on detecting volumetric changes within the ear canal, such activity-related deformations can overwhelm the signal, thereby obscuring vein pulsations and reducing the accuracy of heart rate detection. Figure 1(a)-(b) shows the heartbeat waveform detected by APG in the absence and presence of body motions.

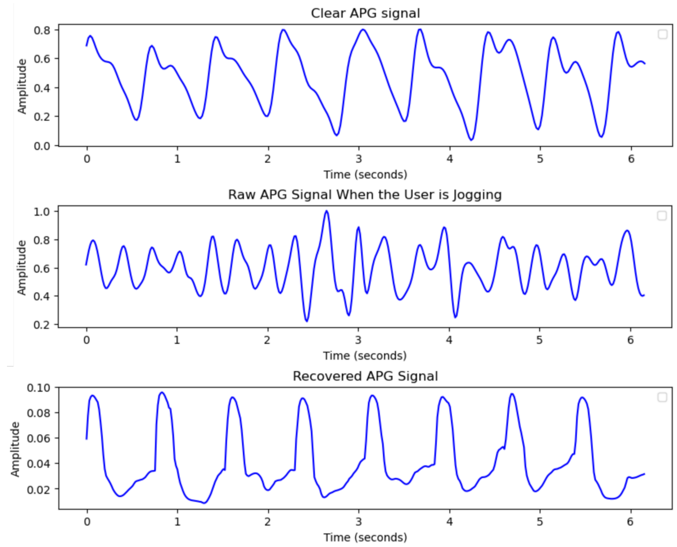

Fig. 1. (a): APG waveform in the absence of body motions; (b): APG waveform in the presence of body motions; (c): Recovered APG waveform.

In this paper, we present CLEAR-APG, a mobile system designed for enhancing APG's robustness to motion artifacts, making continuous heart rate monitoring more accessible. To counteract motion-induced distortions, our basic idea is to explore other information about the body's movement through space. Most commercial earbuds include built-in gyroscope [8], which could serve as a motion reference to aid in APG denoising. However, APG and gyroscope signals operate in different domains: gyroscopes capture three-axis earphone orientation variation, whereas APG measures acoustic reflections. These signals cannot be directly subtracted to remove noise.

In CLEAR-APG, we take gyroscope data as a clue to denoise APG. Inspired by the target speaker extraction (TSE) methods in cocktail party problem [9], we propose to use two separate encoder that extract the signal features from gyroscope readings and microphone recordings and align their features in the high-dimensional space for heartbeat signal denoising. Figure 1(b) and (c) compares the APG signal before and after applying our multi-modal denoising algorithm.

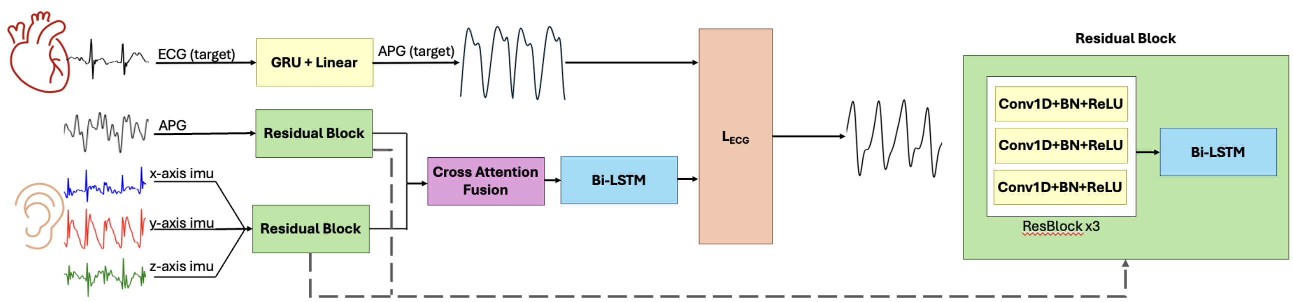

Fig. 2. System overview of CLEAR-APG.

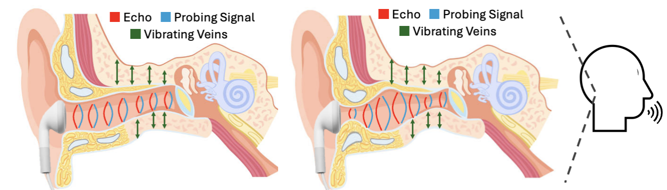

Fig. 3. (Left): APG workflow. (Right): Jaw movements can cause significant deformation of the ear canal, often exceeding the subtle changes caused by blood flow, leading to noisy APG recordings.

In summary, this paper makes the following contributions: (i): We identify a new opportunity to denoise APG signals in the presence of body motion artifacts. (ii): We seize this opportunity by proposing a novel multi-modal solution to enable reliable heart rate monitoring with earphones in unconstrained environment. (ii): We implement our design on earphones and conduct comprehensive experiment across 14 volunteers and six activities. The results show that our solution achieves an consistently low heart rate error rate across all six activities, with an average error rate of 4.01%.

II. Related Works

Electrocardiography (ECG) captures the heart's electrical activity through electrodes placed on the body. It is the gold standard for accuracy but often requires wired setups and professional assistance, limiting its use in continuous, everyday heart rate monitoring [10], [11]. **Photoplethysmography (PPG)** measures blood volume changes using light sensors, typically worn on the wrist, finger, or ear [12]. While common in consumer wearables, PPG is affected by skin tone, sensor quality, and motion artifacts [5]. Additionally, its high power consumption—mainly from LED usage—challenges battery life [13]. **Contactless methods** such as mmWave Radar [14] or Wi-Fi [15] sensing are suited for users with skin sensitivities. They however often requires users to remain still, making them prone to motion artifacts.

III. Design

A. APG at a Glance

APG [7] is a contactless heart rate monitoring approach. As shown in Figure 3(a), it works by emitting a set of ultrasonic probing signal on different frequency band into the ear canal and recording the reflections. Since the ear canal deforms with the vein pulsations inside the ear canal, the echos' waveform will change accordingly and give insight on blood flow dynamics. Figure 1(a) shows the APG waveform.

Challenges. Although APG offers the benefit of contactless sensing, it is highly susceptible to anatomical differences and distortions caused by motion. Specifically, the ear canal not only deforms with the vein pulsations, it also subtly changes during routine activities such as jaw movements (e.g., speaking, chewing), head motions (e.g., tilting up, down, or sideways), facial expressions (e.g., smiling), and general body movements (e.g., walking, jogging, or lifting weights).

As illustrated by Figure 3(b), these body-induced vibrations cause more pronounced structural changes in the ear canal, overwhelming the subtle variations produced by blood flow. As an example, Figure 1(b) shows the APG waveform in the presence of body motions. We can hardly tell the heartbeat cycles, which makes it difficult to extract meaningful cardiovascular metrics. Although APG [7] also employs a blind-source separation-based approach to extract heartbeat waveforms from motion interference, its effectiveness is limited to repetitive patterns, making it less suitable for daily activities.

B. System overview

Motivated by the success of Target Sound Extraction (TSE) in speech enhancement, we propose CLEAR-APG, an end-to-end, cross-modality, learning-based denoiser for robust heart rate monitoring. As shown in Figure 2, the multi-modal denoiser takes two input channels – the noisy acoustic APG signal and a synchronized gyroscope signal. Since the gyroscope captures head and jaw motion, it serves as a proxy for estimating motion-induced distortions in the APG. We leverage this auxiliary motion data through a cross-attention-based neural network that transforms the noisy APG into a cleaner version, better representing true cardiovascular activity. For supervision, we utilize an ECG chest strap as the noise-free reference. To tackle the modality gap between ECG and APG, we introduce an ECG-to-APG translator that synthesizes clean APG signals directly from the corresponding ECG.

C. Multi-Modal Denoiser

The multi-modal denoiser comprises two modality-specific encoders, a cross-attention fusion layer, and a Bi-LSTM layer. The model begins by encoding the gyroscope (IMU) and noisy APG signals separately using two dedicated LSTM-based encoders. This separation allows each encoder to learn modality-specific features, enabling better representation learning and improving robustness against overfitting. Each encoder in-

Fig. 4. (left) Earphone data collection platform. (Middle) The participant is at rest. (Right) The participant is walking. Groundtruth ECG chest strap collects data simultaneously with the earbuds.

cludes a stack of 1D convolutional layers, capturing both local and long-range temporal patterns.

Afterwards, we employ a cross-attention mechanism to guide the fusion process. The encoded IMU features act as contextual cues to inform attention over the APG features, allowing the model to selectively emphasize motion-induced noise patterns. The attended representations are concatenated and passed through a linear layer to generate a compact 256-dimensional fused embedding. The fused embedding is then fed into a Bi-LSTM layer to model temporal dependencies and refine signal reconstruction. This enables the network to leverage sequential context from both forward and backward directions, enhancing denoising performance across time.

D. ECG to APG Translator

Since we have no APG ground-truth while the participant is moving, we use chest strip-based ECG sensor to collect reliable ECG signal during body motion and further design a lightweight ECG-to-APG translator. The NN-based ECG-to-APG translator consists of a Gated Recurrent Unit (GRU) with a hidden size of 64 and 3 layers, and a linear layer. It learns a mapping from ECG waveforms to APG signals, preserving key temporal features while enabling efficient signal transformation. This minimalist design avoids overfitting, generalizes well across cardiac conditions, and serves as a stable source of clean APG-like supervision for training the denoiser.

E. Combine them together

The predicted clean APG from the denoiser is supervised using the reference APG generated by the ECG-to-APG translator. We applied a custom loss function between two signals, combining derivative loss with standard mean squared error (MSE) loss. The derivative loss ensures signal smoothness by calculating the differences in first and second-order gradients, preventing abrupt fluctuations. The MSE loss minimizes discrepancies between the model-generated signal and the target signal at data points, avoiding prediction drift.

IV. Evaluation

The evaluation is structured into two key phases: ECG-to-APG translation and End-to-End APG denoising. We assess the performance of these two parts using mean absolute percentage error on heart rate. This study was approved by the Institutional Review Board (IRB).

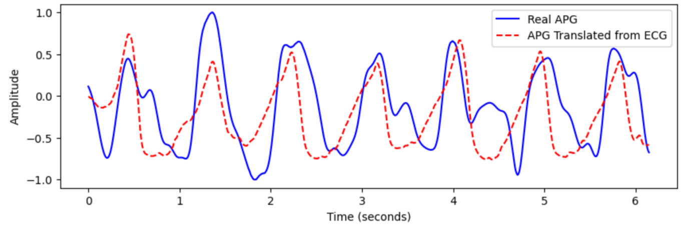

Fig. 5. A snapshot of APG signal translated from ECG signal.

A. System Set Up

To evaluate CLEAR-APG, we first build a data collection platform integrating multiple sensor modalities. For ECG data, we utilize the Polar H10 chest band [16], while gyroscope data is captured using Open Earable [17]. For the APG, we use an earphone equipped with a speaker and feedback microphone, similar to the one used in APG [7]. Each of these hardware components relied on their individual platforms for data acquisition. To synchronize data collection across these modalities, we developed a Node.js server that managed communication and recorded metadata, such as participant IDs and activity labels. Figure 4 shows the data collection process.

14 participants were invited to perform six activities: resting, talking, simulated chewing, simulated weightlifting, walking, and jogging. The APG and ECG data were downsampled to match up with the 50 Hz sampling frequency of the IMU. A low-pass anti-aliasing filter (cutoff frequency of 25 Hz) was applied before downsampling to prevent frequency aliasing and to preserve low-frequency heart rate signals. We also recruited another 12 participants for ECG-to-APG translator training and testing. These participants remained still while ECG and APG data were collected. We collect these biomarkers for each participant 10 to 20 periods, with each period lasting 10 seconds. Each audio period is further segmented into 5-second chunks for training, resulting in a total of 650 chunks. We employ a 5-fold cross validation to evaluate our system.

B. Experiment Results

1) ECG to APG Translation Model: In total, we use 120 chunks of paired ECG and APG data (lasting for 20 minutes) for evaluation. We show a snapshot of the APG signal translated from ECG signal in Figure 5. Although the translated APG waveform (from ECG) is not completely same as the real APG waveform, it maintains a high similar coherence with the real APG signal, showing a very pronounced periodicity. We validated the model using a test set comprising ten groups of data, each containing ten seconds of APG and ECG. The results showed that the HR calculated from translated APG had an average percentage error as low as 0.38% compared to the HR from the real APG, which confirms the efficacy of our ECG-to-APG translator.

2) End-to-End Denoising Model: We use the percentage error between the HR calculated from the original APG and the HR calculated from the denoised APG signal as a criterion to measure the model's performance. Due to the

TABLE I
AVERAGE HR PERCENTAGE ERROR ACROSS DIFFERENT ACTIVITIES.

(%)	Rest	Talk	Chew	Weightlift	Walk	Jog	Average
Raw APG	5.92	21.40	33.94	23.47	32.65	23.65	23.50
Recovered APG	2.36	3.34	4.67	6.43	4.02	3.25	4.01

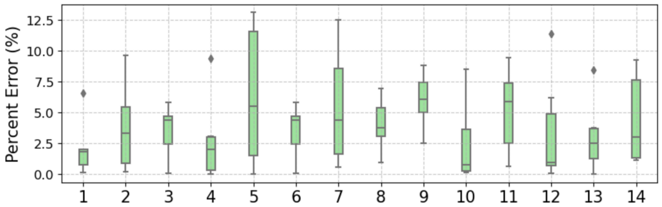

Fig. 6. Percent error distribution by volunteer

limited number of volunteers (14 in total), We apply cross-validation to train the model on a set of volunteers and test the model on the remaining ones. We repeat this process and obtain the HR error rate for all 14 volunteers to fully assess the model's performance across individuals. Figure 6 shows the distribution of heart rate error rate on each volunteer. It can be seen that most of the errors are below 10%. Then, we average the HR calculated for each scenario across the volunteers and compared the average percentage error between the HR calculated from the original APG and the HR calculated from the denoised APG. Table I shows the differences between them. We observe that the average HR error from the original APG was as high as 23.50%, while the error from the recovered APG decreased to 4.01%. Specifically, the error is smallest at rest, at 2.36%, and largest at weightlift, at 6.43%. The high error during weightlifting may be due to irregular limb movements, which poses a challenge to our denoising model. Nevertheless, all errors are still well below the industry-required tolerance range of 10% [18].

V. Discussion

Although our evaluation shows that CLEAR-APG significantly improves the robustness of APG-based HR monitoring, it is worth acknowledging that the current work still has some presenting challenges leave for future research:

• **Small dataset size**. Our current evaluation is restricted to a small number of volunteers. In the future, we plan to validate on larger datasets to better assess cross-user performance.

• **ECG-to-APG translator**. Current CLEAR-APG utilizes a suboptimal clean APG generated by the ECG-to-APG translator as groundtruth reference, which may introduce intrinsic errors (average HR error of 0.38%). Consequently, the denoiser may learn these discrepancies, potentially amplifying inaccuracies in high-variability activities. To address this, future work could explore direct collection of motion-free APG ground-truth or refine the translator with larger datasets.

• **Data augmentation**. Furthermore, modeling noise patterns – such as periodic deformations from walking or vibrations during speaking could offer an avenue for simulation-based data augmentation. By generating synthetic noisy signals from clean ground truth, we can expand the dataset and enhance model robustness.

VI. Conclusion

We have presented CLEAR-APG, a multi-modal approach to APG denoising. Inspired by target sound extraction in speech processing domain, we leverage gyroscope readings as a motion clue to guide the model to gradually remove the impact of body motions on raw APG signals, ensuring reliable heart rate monitoring in different daily activities setups.

Acknowledgment

This work is supported by the National Science Foundation under Grant No.2337537.

References

[1] R. Albaghli and K. M. Anderson, "A vision for heart rate health through wearables," ser. UbiComp '16.

[2] J. P. L. Leenen, C. Leerentveld, J. D. van Dijk, H. L. van Westreenen, L. Schoonhoven, and G. A. Patijn, "Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: Systematic review," *J Med Internet Res*, 2020.

[3] "Apple watch," https://www.apple.com/watch/?afid=p238%7CsNZgeoZeS-dc_mtid_1870765e38482_pcrid_661421631506_pgrid_99322576784_pntwk_g_pchan__pexid__&cid=aos-us-kwgo-watch--slid---product-.

[4] D. Biswas, N. Simões-Capela, C. Van Hoof, and N. Van Helleputte, "Heart rate estimation from wrist-worn photoplethysmography: A review," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6560–6570, 2019.

[5] B. Bent, B. A. Goldstein, W. A. Kibbe, and J. P. Dunn, "Investigating sources of inaccuracy in wearable optical heart rate sensors," *npj Digital Medicine*, vol. 3, no. 1, p. 18, Feb 2020. [Online]. Available: https://doi.org/10.1038/s41746-020-0226-6

[6] A. L. Ries, L. M. Prewitt, and J. J. Johnson, "Skin color and ear oximetry," *Chest*, vol. 96, no. 2, pp. 287–290, 1989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0012369215459161

[7] X. Fan, D. Pearl, R. Howard, L. Shangguan, and T. Thormundsson, "Apg: Audioplethysmography for cardiac monitoring in hearables," ser. ACM MobiCom '23.

[8] "Airpods," https://www.apple.com/airpods/.

[9] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 09 1953. [Online]. Available: https://doi.org/10.1121/1.1907229

[10] S. Meek and F. Morris, "ABC of clinical electrocardiography.introduction. I-Leads, rate, rhythm, and cardiac axis," *BMJ*, vol. 324, no. 7334, pp. 415–418, Feb. 2002.

[11] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "Ptb-xl, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 1, p. 154, May 2020. [Online]. Available: https://doi.org/10.1038/s41597-020-0495-6

[12] A. Challoner, "Vj.(1979). photoelectric plethysmography for estimating cutaneous blood flow," *ROLFE, P. Academic Press London. New York. San fransisco*, pp. 125–151.

[13] A. Burt, "Need clinical-grade ppg from your wearable? sometimes it pays not to shine a light on a problem," https://www.analog.com/en/resources/technical-articles/need-clinical-grade-ppg-from-your-wearable.html, 2022.

[14] W. Xia, Y. Li, and S. Dong, "Radar-based high-accuracy cardiac activity sensing," *IEEE Transactions on Instrumentation and Measurement*, 2021.

[15] X. Wang, C. Yang, and S. Mao, "On csi-based vital sign monitoring using commodity wifi," *ACM Transactions on Computing for Healthcare*, 2020.

[16] "Polar h10," https://www.polar.com/us-en/sensors/h10-heart-rate-sensor?srsltid=AfmBOoqrxh71X6GsAHp5gr-_miJFBcoVc9sFGQGqxiw6Co95Dn5tQVPa.

[17] T. Röddiger, M. Küttner, P. Lepold, T. King, D. Moschina, O. Bagge, J. A. Paradiso, C. Clarke, and M. Beigl, "Openearable 2.0: Open-source earphone platform for physiological ear sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 9, no. 1, pp. 1–33, 2025.

[18] A. for the Advancement of Medical Instrumentation, *Cardiac Monitors, Heart Rate Meters, and Alarms*. Arlington, VA: Association for the Advancement of Medical Instrumentation, 2002, aNSI/AAMI EC13:2002. [Online]. Available: https://www.aami.org/