Deep-Learning Based Segmentation of In-Ear Cardiac Sounds

Jordan Waters, Jake Stuchbury-Wass, Yang Liu, Kayla-Jade Butkow and Cecilia Mascolo
University of Cambridge, UK
jw2190@cantab.ac.uk, {js2372, yl868, kjb85, cm542}@cam.ac.uk

***Abstract*— Cardiovascular disease is a leading cause of death worldwide. Auscultation, a common diagnostic method, involves listening to heart sounds to detect valve contraction irregularities, with heart sound segmentation (i.e., identifying heartbeat phases) being a crucial first step. Due to the high level of expertise required for traditional auscultation, previous work has automated segmentation using digital sounds from stethoscopes or phonocardiographs; however, these methods depend on specialist medical devices, limiting continuous and wide usage. Earable devices now offer continuous heart sound capture via in-ear microphones (IEM) on Active Noise Cancellation (ANC) earphones, opening new possibilities for portable, continuous, out-of-hospital heart sound segmentation. However, techniques developed for phonocardiogram (PCG) signals are not directly applicable because of the distinct differences in signal characteristics between IEM and PCG signals. In this work, we are the first to demonstrate the potential of using in-ear cardiac sounds for heart sound segmentation. We begin by analysing the temporal and frequency differences between IEM and PCG signals, then introduce a U-Net deep learning model tailored for in-ear heart sound segmentation. Additionally, we propose a more stringent evaluation method for segmentation accuracy and use this to evaluate our method and the baselines. We collected data from 11 participants, and our model achieved 84% accuracy, outperforming established baselines.**

I. Introduction

Cardiovascular disease is a major cause of premature mortality and rising healthcare costs. This disease is only growing in severity, with the number of global cases doubling between 1990 and 2019 from 271 million to 523 million [1]. One method of diagnosing cardiovascular disease is through auscultation of the heart, the process of a medical professional listening to the sounds produced as it beats, typically with a stethoscope [2]. For example, abnormal heart sounds and irregular timings between heart valve contractions can help diagnose heart murmurs and valve disease [2].

One of the first steps of auscultation is identifying phases of the heartbeat based on the heart sounds. As shown in Figure 1, a heartbeat can be divided into 4 main phases: S1 (closing of atrioventricular valves), systole (ventricular contraction/ejection), S2 (closure of semilunar valves), and diastole (ventricular filling) [3], [4]. Once these phases are identified, abnormal heart sounds can be classified with a possible diagnosis made. For example, mitral stenosis can be identified as a low rumbling sound in the diastolic phase, whereas aortic stenosis is a harsh sounding high-pitched noise in the systolic phase [3]. Therefore, heart sound segmentation, identifying the timings of these four main phases of heart sounds, is an important initial step for auscultation of the heart, which provides critical insights into cardiac function and potential abnormalities.

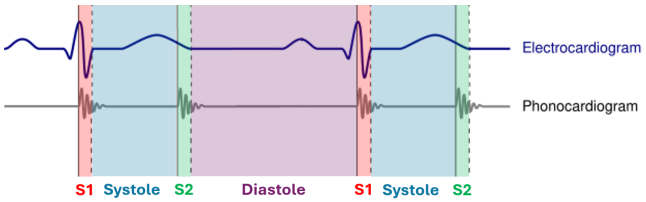

Fig. 1: The 4 fundamental phases of a heartbeat in a PCG and electrocardiogram (ECG) recording.

However, auscultation is a difficult skill to acquire among professionals, with only 20% of medical interns able to detect heart conditions effectively [5] and it is often unavailable in some regions of the world, where the patient-to-doctor ratio sometimes reaches 50000:1 [6]. For this reason, various algorithms for automatic heart sound segmentation have been proposed over the past decades. Work by Lhener et al. which dates back to 1987, used signal processing techniques on a combination of PCG, ECG and carotid pulse signals [7]. After decades of research, segmentation methods can now be grouped into 3 main categories: envelope-based methods [8], [9], [10], hidden Markov models [11], [12], [13], and machine learning models [14], [15], [16], [17]. Schmidt et al. proposed hidden semi-Markov models (HSMMs) as an extension to the hidden Markov model, reaching sensitivities of up to 99.1% [18]. Springer et al. built upon this work, achieving accuracies of 95.6% [19]. However, both the manual and traditional automatic methods require hospital resources. Data must be collected by specialist medical devices which limits continuous usage, or clinicians are required. These requirements of hospital resources reduce the accessibility of this segmentation technology, and therefore its rate of adoption.

Recently, the wearable device industry is rapidly growing, with 1.1 billion wearable devices in use in 2022, up from 929 million in 2021 [20]. The growing use of wearable devices is beneficial to the medical industry as it is expected to save around $200 billion in costs over the next 25 years and rapidly decrease the interaction time between clinicians and patients [21]. Additionally, ANC earbuds have become ubiquitous in daily life. These earbuds contain a microphone facing inside the ear canal (IEM) which achieves noise-cancellation functionalities. Lately, it has been proven that heart rate can be measured accurately using inwards facing

microphones [22], with other work proving that repurposed ANC earphones can record clear and reliable heartbeat sounds [23], [24]. In addition, the nature of the mechanisms of production for in-ear cardiac sounds has been researched, with a study showing they contain components from both conducted sounds and pulse waves around the ear [25].

This suggests that earbuds can be the next frontier for automatic auscultation of the heart. In this work, we take the first step toward providing portable, continuous, and out-of-hospital heart sound segmentation capabilities using in-ear microphones on earphones. This is of significant value to the field of auscultation for the following reasons:

- **Increased accessibility:** The use of commodity devices, i.e., ANC earbuds, already owned by large portions of the population enables wider access to cardiac auscultation, especially in out-of-hospital and everyday settings.
- **Continuous monitoring:** Earphones are portable, non-invasive and user-friendly, and can be used comfortably for long periods. This enables longitudinal monitoring of cardiac auscultation.
- **No clinician required:** Unlike an electronic stethoscope or phonocardiograph, positioning and operation of earphones is simple. This means that it can be done at home, thus facilitating telemedicine.

Although existing studies enable heart sound segmentation on PCG signals [14], [19], the techniques developed for these signals are not directly applicable due to the distinct differences in signal characteristics between in-ear audio and PCG signals (as detailed in Section III-A). In this work, we investigate the feasibility of using audio collected inside the ear canal for heart sound segmentation. We also determine the properties of the IEM signal for cardiac auscultation and segment the IEM cardiac sounds with high accuracy. Our contributions are summarized as follows:

- **Characterization of IEM signals:** We identified: **(i)** Potential timing differences between the IEM and associated PCG signal. **(ii)** Frequency differences between the IEM and PCG signals.
- **IEM heart sound segmentation pipeline:** We devised a pipeline for segmenting IEM signals, using a deep convolutional neural network (CNN), i.e., U-Net model. We compared its performance to previous state-of-the-art models tailored for PCG signal segmentation (i.e., Renna et al.'s [14] and Springer et al.'s models [19]). Evaluations on data collected from 14 participants demonstrated the superior performance of our pipeline.
- **Segmentation evaluation:** We developed an evaluation technique to allow strict comparison between models run on IEM signals.

II. User Study

1) Devices: While in-ear microphones are now common in ANC earbuds for active noise cancellation, no commercial earbuds provide access to the raw data. Figure 2 illustrates our developed hardware used for data collection. A Knowles SPU1410LR5H-QB microphone, which has a flat frequency response between 10Hz and 10kHz, was embedded into a 3D-printed earbud and used as the IEM. The IEM and PCG recordings were collected simultaneously in our experiments, using the IEM setup shown in Figure 2(a) and a Thinklabs One digital stethoscope [26] shown in Figure 2(b). An appropriately sized ear tip was selected for each participant based on their ear canal shape to ensure a tight seal and good occlusion.

Fig. 2: Data collection devices: (a) Earbuds, and (b) Thinklabs One digital stethoscope.

2) Data collection: We collected data from 11 participants (5 male and 6 female) with an average age of 24.9 ± 4.6, in a study approved by the University of Cambridge Department of Computer Science and Technology Ethics Committee. A summary of the collected data is provided in Table I. The data collection involved recording participant responses under three different background noise conditions: silence, 60 dB, and 70 dB. During each session, subjects sat on a chair and remained stationary. The duration varied, with the silent condition lasting 5 minutes and the noisy conditions lasting 2 minutes each. The number of participants differed across conditions, with 11 in the silent condition, 9 in the 60 dB condition, and 8 in the 70 dB condition. During data collection, heart sounds were collected by the ground truth e-stethoscope. The average duration of each heart sound phase, as measured by the e-stethoscope's PCG, is shown in Table II. Simultaneously, audio was recorded in the participants' ears using the IEMs in the earbuds. The participant tapped their upper body several times at the beginning of each data collection session to synchronize the recordings from the two devices. This generated visible peaks that appeared in both recordings simultaneously, facilitating their alignment.

TABLE I: Summary of data collection.

Test Case	Duration (minutes)	Number of Participants
Silence	5	11
60dB	2	9
70dB	2	8

TABLE II: Durations of each heart sound phase from PCG signals across the entire collected dataset, presented as mean $\pm$ sd (min $-$ max).

Heart Sound	Duration (ms)
S1	124±13 (19-179)
S2	103±12 (39-159)
Systole	180±32 (39-279)
Diastole	425±103 (159-719)

III. Methods

In this section, we present our methodology for heart sound segmentation of IEM signals.

A. Signal characterization

1) IEM signal: When the ear canal is blocked (or occluded), there is an increase in impedance inside the ear canal. This causes low-frequency sounds inside the ear canal to be amplified, and higher frequency sounds to be attenuated [22]. Specifically, this affects bone-conducted sounds: sounds that are propagated by bones to the inner ear, causing vibration in the ear canal walls. Among the most relevant of these bone-conducted sounds are sounds related to heart activity [22]. By placing a microphone inside the occluded ear canal, these heart signals can be captured. Recent work identified that these signals contain a combination of compressive sound waves (heart sounds) originating from the chest and pulse waves from arterial expansion near the ear canal [27].

The heart signals captured from the IEMs are shown in Figure 3(b). Figure 3(a) shows the simultaneously captured heart signals from the digital stethoscope. By comparing Figure 3(a) and Figure 3(b), it is evident that due to the compressive component, the cardiac signal contains clear heart cycles, with evidence of an S1 peak, S2 peak, and systolic and diastolic periods between these. These cycles also align with those in the PCG signal.

2) Frequency content variation between PCG and IEM signals: However, in agreement with Christofferso et al. [27], it is also evident that the in-ear cardiac signal has different characteristics to the PCG signal. We first analysed the variation in frequency content by plotting the amplitude of the IEM and PCG signals across frequency bands, as shown in Figure 4(a) and Figure 4(b), respectively. Due to the occlusion effect, the IEM heart signal captured inside the ear canal has much stronger lower frequency components than the PCG. We see that for the IEM signal, all frequency content above 30 Hz is weak, whereas for the PCG signal, there is strong frequency content up to 250 Hz.

3) Time offset between PCG and IEM signals: After aligning the PCG and IEM signals using the synchronisation activity, we observed that the peaks in the IEM signal corresponding to the S1 and S2 sounds exhibited a time offset relative to the PCG peaks, as shown in Figure 3(c). These locations are identified as the onset of the first high-energy part of the IEM signal after the PCG S1 or S2 onset. We further explored the durations of these offsets to identify whether their source was experimental or due to human physiology. Across all test cases, the average delay for the right channel was 97.1 ± 31.2 ms, compared to 92.3 ± 23.7 ms for the left channel (using format mean$\pm$stdev). Figure 4(c) suggests that the delays are different for each participant, while still being in the range 50-150 ms. We expect that the delay depends on a multitude of factors, including the participants' cardiac output at the time of recording.

This consistency in channel delay implies that the left ear experiences the signal before the right. This paired with the variability in delay length implies that the source is not experimental, but rather due to the changing physiology and body composition of the participants. The speed of sound in blood is 1570 m/s [28], [29] and is even faster in bone, meaning that it takes 0.6 ms for heart sounds to travel one meter of blood vessels. As the delays observed are much greater than this value, we hypothesise that rather than detecting sounds from the heart valves themselves, the IEMs are detecting pulse pressure waves due to blood pumping through the vessels. This is supported by Gårdbæk et al.'s work [30]. Recent work found that the in-ear cardiac signal appears to be composed of both a bone-conducted and pulse wave components [25]. Our analysis indicates that the pulse wave component is dominant in a set-up with high ear canal occlusion.

Our characterisation of the IEM heart signal has shown that it has a significantly different frequency spectrum and timing characteristics to the PCG signal. As such, PCG-based segmentation approaches cannot directly be applied to the IEM signal. Therefore, we develop new approaches to segmenting the IEM signal, as detailed in the coming sections.

B. S1 Synchronisation

To remove the time offset between PCG and IEM signals for model training, we further synchronised the IEM and PCG S1 phases. To achieve this, we identified the peak in the PCG signal corresponding to the start of an S1 phase. The corresponding delayed peaks in the IEM channels were then detected and shifted to align with the PCG peak. Figure 3(d) shows the results of synchronising the S1 peaks, indicating the perfect alignment of the IEM and PCG signals once the offsets have been removed.

C. Segmentation

1) Feature vector generation: We first divide the IEM signals into windows of 1.28 seconds, with an overlap of 1.12 seconds between consecutive windows. This process generates feature vectors consisting of 64 data points, as explained below, with an 8-point stride, to be used as model input.

We generate a feature vector for each window of input, following the approaches of Springer et al. [19] and Renna et al.'s work [14]. Specifically, we extract a four-dimensional feature vector from each channel of IEM signals, comprising the following components:

- Homomorphic envelope: Obtained by exponentiating the low-pass-filtered natural logarithm of the IEM signal

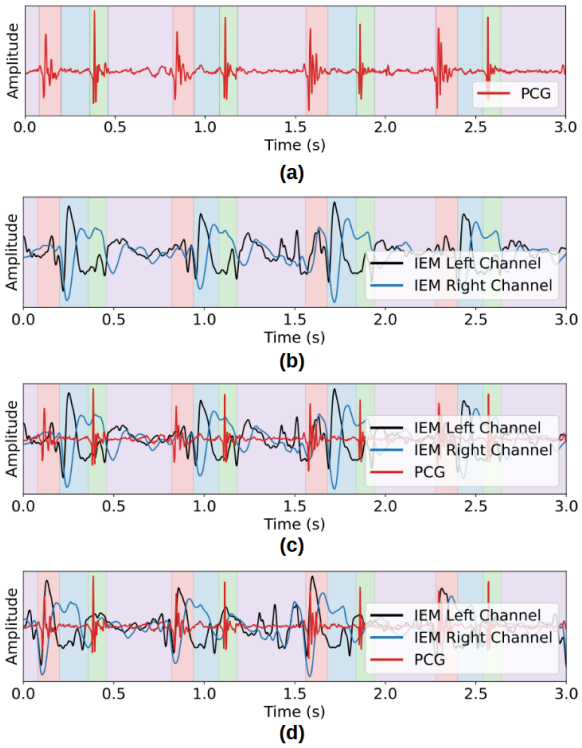

Fig. 3: PCG and IEM signals with segmentation regions annotated. S1 (Red), Systole (Blue), S2 (Green) and Diastole (Purple). (a) PCG signal. (b) Normalised, time-synchronised IEM Left and Right Channels. (c) Normalised, time-synchronised IEM Left and Right channels with PCG. (d) Normalised, S1-synchronised IEM Left and Right Channels with PCG.

to extract its envelope [31]. We use a low-pass filter with a 25 Hz cutoff, which effectively preserves critical low-frequency heart signals in the IEM recordings.

- Hilbert envelope: Computed as the absolute value of the Hilbert transform, which preserves only the positive-frequency components of a signal by extracting its analytic signal [32], [33].
- Wavelet envelope: Extracted using a level-5 Daubechies wavelet, selected based on prior work [8], [34].
- Power Spectral Density (PSD) envelope: Computed via the short-time Fourier transform (STFT) after applying a Hamming window [19].

Both signal channels were then concatenated into a single feature vector, which served as the model input.

2) Segmentation model: To segment the IEM signal, we proposed a CNN-based model, which follows a deep CNN architecture [14] designed to process per-channel feature vectors from audio data. As shown in Fig. 5, the model consists of multiple convolutional layers with ReLU activation, downsampling through max pooling, upsampling layers, and skip connections. The input consists of concatenated feature vectors from the left and right IEM channels, and the output is a sequence of segmentation labels corresponding to different heart phases.

As discussed in Renna et al. [14], this deep CNN architecture effectively segments the IEM signal by leveraging a large receptive field. Neighbouring data points significantly influence the probability of the current point belonging to one of the four heart phase states. This allows the model to accurately capture the temporal evolution of the signal and corresponding heart sounds. By doubling the size of the input feature vector, i.e., incorporating both left and right channels simultaneously, the model gains additional robustness, enabling it to consider both channels when determining heart sound labels.

Following Renna et al. [14], the predicted states undergo a sequential temporal check that corrects invalid sequences. Since a healthy heart always follows the pattern: S1 $\rightarrow$ systole $\rightarrow$ S2 $\rightarrow$ diastole $\rightarrow$ S1..., the check removes heart phase sequences that do not follow this pattern.

In addition, training was designed to encourage the model to learn distinct features from each channel. The process involved:

1) Train the CNN on the feature vector containing the left and right features.
2) Replace all left channel features with zeros and repeat training.
3) Replace all right channel features with zeros and repeat training.

This training method forced the model to develop separate feature maps for each channel, capturing unique characteristics independently. This method therefore ensures that high segmentation performance can be obtained even in the presence of poor-quality signals in one channel.

IV. Evaluation

A. Training and testing

The PCG heart sound labels produced by the Springer HSMM [19] were used to train and test all models tasked with segmenting IEM audio. The Springer HSMM was used due to its accuracies of 95.6% segmenting PCG audio [19]. The primary validation method was Leave One Subject Out cross validation (LOSO) over the 11 participants. In each case the model was trained on 10 recordings, with the final used for testing, then performance was averaged across all 11 trials.

B. Evaluation metrics

In typical heart sound segmentation, PCG sounds are segmented using ECG as a reference to evaluate performance. A specialist or ECG segmentation algorithm is used to identify the ground truth labels. However, evaluating heart sound labels produced by a PCG against those produced by an ECG is a non-trivial task. This is due to the fact that ECG peaks, in the electrical domain, do not match PCG peaks, recorded in the sound domain (Figure 1). Therefore, evaluation of labels is typically done by creating a "tolerance", with which the labelled sound must fall within relative to the ECG label. Springer et al. labelled an S1 sound as a true positive if it lay within 100ms of the ECG's R-peak [19].

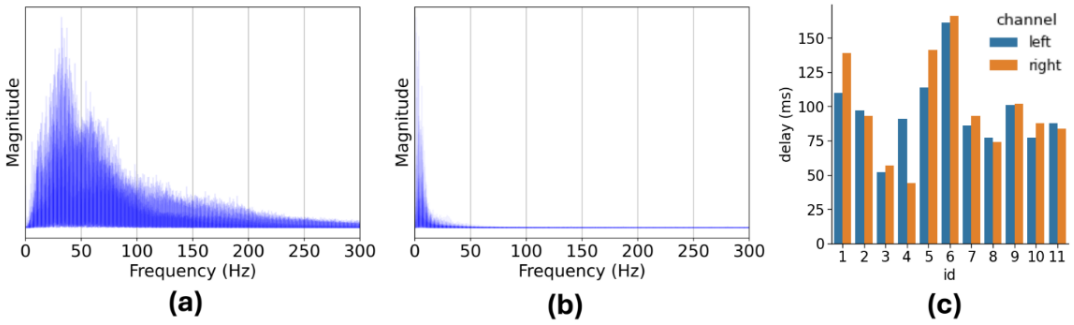

Fig. 4: Characterisation of the in-ear cardiac signal (a) Frequency spectrum of PCG signals and (b) IEM signals. (c) Delay in milliseconds between the IEM signals and the corresponding PCG signals per participant.

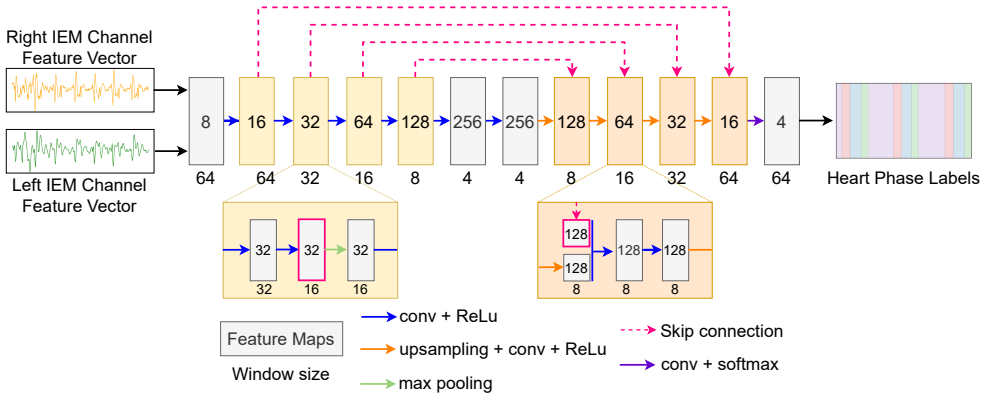

Fig. 5: Segmentation model architecture. The model's input consists of per-channel feature vectors, and the output is the segmentation labels.

TABLE III: Accuracy and F1 scores of our model compared to the Renna CNN [14] and the Springer HSMM [19] (Average $\pm$ 1 standard deviation). The best results are shown in bold.

Model	Trained On	Tested On	Ac	P+	Se	F1 (S1)	F1 (S2)	F1
Springer	PCG	IEM Left	0.7 ± 0.07	0.58 ± 0.09	0.57 ± 0.09	0.76 ± 0.06	0.62 ± 0.09	0.58 ± 0.09
Springer	IEM	IEM Left	0.53 ± 0.1	0.41 ± 0.12	0.4 ± 0.11	0.62 ± 0.1	0.54 ± 0.05	0.41 ± 0.11
Renna	IEM	IEM Left	0.79 ± 0.2	0.75 ± 0.2	0.75 ± 0.2	0.84 ± 0.16	0.77 ± 0.15	0.74 ± 0.21
Our Model	IEM	IEM Left	0.81 ± 0.17	0.79 ± 0.14	0.78 ± 0.18	0.87 ± 0.1	0.8 ± 0.15	0.77 ± 0.19
Our Model	IEM	IEM Right	0.76 ± 0.21	0.72 ± 0.21	0.71 ± 0.22	0.84 ± 0.15	0.73 ± 0.15	0.7 ± 0.23
Our Model	IEM	IEM Both	**0.83±0.19**	**0.82±0.14**	**0.8±0.18**	**0.88±0.12**	**0.81±0.14**	**0.79±0.21**

Se denotes sensitivity (aka recall), $P+$ is positive predictivity (aka precision), Ac is accuracy, $F1(S1)$ is F1 score for S1 sounds, $F1(S2)$ is F1 score for S2 sounds, $F1$ is F1 score for all heart sound stages.

In our work, evaluation of the IEM signal was done by directly comparing it to ground truth labels (on a sample-by-sample basis) produced by a PCG signal rather than ECG to obtain more accurate labels. Each predicted sample was compared to the actual heart sound label which allows us to separate every sample into the following categories: True Positive, False Positive, False Negative and True Negative. After this classification, the following metrics were calculated:

- **Accuracy (Ac):** The proportion of true predictions.
- **Positive predictivity ($\mathbf{P_+}$):** Represents how many positive predictions are true, otherwise known as precision.
- **Sensitivity (Se):** Represents how well true positives are identified, otherwise known as recall.
- **F1 Score (F1):** The harmonic mean of sensitivity and positive predictivity. F1 score is used in this context as the systole and diastole have much longer durations than S1 and S2, which can lead to misleading accuracy scores. Increasing positive predictivity often leads to decreasing sensitivity, so the F1 score combines these two scores to indicate when both of these metrics have improved. Due to this, we use F1 score as our primary evaluation metric.

Deriving these metrics for every sample rather than over relatively large windows allows a much stricter evaluation of the segmentation method compared to the literature.

C. Segmentation results

We present our overall results in Table III which shows that our model outperforms the Renna CNN [14] and the

Springer HSMM [19], with segmentation accuracy of 0.84 ± 0.05 and F1 score of 0.79 ± 0.05. The table also shows F1 specifically for S1 and S2 segmentation, showing that S1 segmentation has better performance than the other regions. We also provide a visualisation of our results in Figure 6 where we compare the ground truth and predicted labels for a segment of in-ear audio. This figure demonstrates that states are rarely missed and predictions are highly accurate, proving the effectiveness of our system. Additionally, in Figure 8 we show the variance in F1 score for each participant, this shows some variance per person with some participants performing better than others, this could be caused by audio quality in both PCG and IEM signals as well as individual body composition.

To evaluate the performance of our model in noisy environments, we conducted a single-tailed t-test with 95% confidence interval. We found that segmentation performance did not decrease at 60 dB and 70 dB of background noise. This is also illustrated in Figure 7, where performance was comparable in all test cases, in fact, a small increase in F1 score was seen in the noisy test cases.

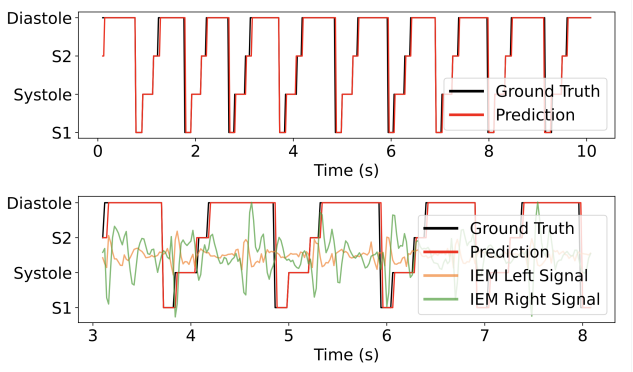

Fig. 6: Segmentation results showing the predicted labels versus the ground truth labels for a segment of in-ear audio.

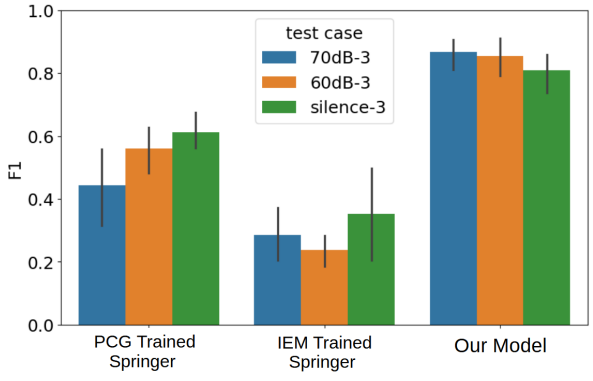

Fig. 7: Background noise results with F1 score in 60dB, 70dB and silent test cases, with error bars representing two standard deviations

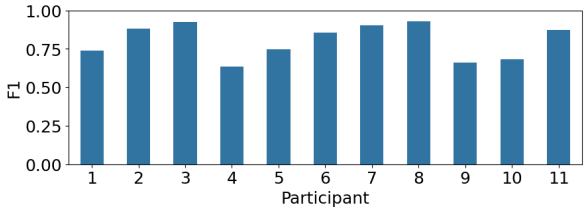

Fig. 8: F1 Score for each participant

V. Discussion

We have proven, for the first time, the feasibility of segmenting in-ear cardiac sounds into the four phases of the cardiac cycle. We achieved high performance with an accuracy of 84% and an F1 score of 79%. This performance is especially high when considering our stricter sample-by-sample evaluation method. This study also considers the nature of the in-ear cardiac sounds, showing substantial delays from chest heart sounds indicative of the in-ear sounds being produced by pulses at the ear from blood movement.

However, our work has several limitations. Firstly, our dataset only includes data from 11 participants. Future work should collect data from more participants with a wider range of ages and demographics to ensure generalisability of our system. Additionally, our work assesses only cardiac sounds from healthy participants. In the future, data should be collected from participants with cardiovascular diseases to ensure that segmentation can still be accurately done in the presence of heart murmurs and other heart sound abnormalities.

Our work takes the first step toward providing portable, continuous, and out-of-hospital heart sound segmentation capabilities using in-ear microphones on earphones. Although work is required to extend our findings to participants with cardiovascular diseases, this work is a promising initial step toward future heart disease detection and longitudinal cardiac health monitoring with telemedicine.

VI. Conclusions

This work has shown, for the first time, that in-ear cardiac sounds can be segmented into the four fundamental heart sound phases with high accuracy. We have outlined a new, stricter evaluation for heart sound segmentation designed for IEM signals when evaluating with ground truth PCG labels. We presented a new deep learning model for segmenting IEM signals, which achieves accuracies of 84% and a 79% F1 Score. The properties of IEM signals have been presented, along with a finding that IEM cardiac signals experience a delay compared to their associated PCG signals. Further research is needed into the cause of these delays and the reason for their variation between participants.

Acknowledgements

This work is supported by ERC through Project 833296 (EAR) and Nokia Bell Labs through a donation.

REFERENCES

[1] G. A. Roth *et al.*, "Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the gbd 2019 study," *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982–3021, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0735109720377755

[2] L. K. Colin Tidy, "Heart auscultation - types of heart sounds and murmurs," Patient, 12 2021. [Online]. Available: https://patient.info/doctor/heart-auscultation

[3] S. Leng, R. S. Tan, K. Chai, C. Wang, D. Ghista, and L. Zhong, "The electronic stethoscope," *Biomedical engineering online*, vol. 14, p. 66, 07 2015.

[4] A. N. M. Joshua D. Pollock, *Physiology, Cardiac Cycle*. Treasure Island (FL): StatPearls Publishing, 2022, [Updated 2022 Oct 3]. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK459327/

[5] S. Mangione, "Cardiac auscultatory skills of physicians-in-training: a comparison of three english-speaking countries," *The American Journal of Medicine*, vol. 110, no. 3, pp. 210–216, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0002934300006732

[6] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, and R. G. Mark, "Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016," in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 609–612.

[7] R. J. Lehner and R. M. Rangayyan, "A three-channel microcomputer system for segmentation and characterization of the phonocardiogram," *IEEE Transactions on Biomedical Engineering*, vol. BME-34, no. 6, pp. 485–489, 1987.

[8] L. Huiying, L. Sakari, and H. Iiro, "A heart sound segmentation algorithm using wavelet decomposition and reconstruction," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No.97CH36136)*, vol. 4, 1997, pp. 1630–1633 vol.4.

[9] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in *Computers in Cardiology 1997*, 1997, pp. 105–108.

[10] A. Moukadem, A. Dieterlen, N. Hueber, and C. Brandt, "A robust heart sounds segmentation module based on s-transform," *Biomedical Signal Processing and Control*, vol. 8, no. 3, pp. 273–281, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809412001280

[11] P. Sedighian, A. W. Subudhi, F. Scalzo, and S. Asgari, "Pediatric heart sound segmentation using hidden markov model," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 5490–5493.

[12] Y.-J. Chung, "Classification of continuous heart sound signals using the ergodic hidden markov model," in *Proceedings of the 3rd Iberian Conference on Pattern Recognition and Image Analysis, Part I*, ser. IbPRIA '07. Berlin, Heidelberg: Springer-Verlag, 2007, p. 563–570.

[13] C. Lima and M. Cardoso, "Phonocardiogram segmentation by using hidden markov models," *Proceedings of the 5th IASTED International Conference on Biomedical Engineering, BioMED 2007*, 02 2007.

[14] F. Renna, J. Oliveira, and M. T. Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2435–2445, 2019.

[15] Y. He, W. Li, W. Zhang, S. Zhang, X. Pi, and H. Liu, "Research on segmentation and classification of heart sound signals based on deep learning," *Applied Sciences*, vol. 11, no. 2, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/2/651

[16] E. Messner, M. Zöhrer, and F. Pernkopf, "Heart sound segmentation—an event detection approach using deep recurrent neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1964–1974, 2018.

[17] Y. Chen, J. Lv, Y. Sun, and B. Jia, "Heart sound segmentation via duration long–short term memory neural network," *Applied Soft Computing*, vol. 95, p. 106540, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494620304798

[18] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden markov model," *Physiological measurement*, vol. 31, no. 4, p. 513, 2010.

[19] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-hsmm-based heart sound segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2016.

[20] "Global connected wearable devices 2016-2022," Statista. [Online]. Available: https://www.statista.com/statistics/487291/global-connected-wearable-devices/statisticContainer

[21] V. Vijayan, J. P. Connolly, J. Condell, N. McKelvey, and P. Gardiner, "Review of wearable devices and data collection considerations for connected health," *Sensors*, vol. 21, no. 16, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/16/5589

[22] K.-J. Butkow, T. Dang, A. Ferlini, D. Ma, and C. Mascolo, "hEARt: Motion-resilient heart rate monitoring with in-ear microphones," in *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 3 2023.

[23] X. Fan, L. Shangguan, S. Rupavatharam, Y. Zhang, J. Xiong, Y. Ma, and R. Howard, "Headfi: bringing intelligence to all headphones," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 147–159.

[24] Y. Cao, C. Cai, A. Yu, F. Li, and J. Luo, "Earace: Empowering versatile acoustic sensing via earable active noise cancellation platform," vol. 7, no. 2, jun 2023. [Online]. Available: https://doi.org/10.1145/3596242

[25] K. Christofferson, S. Bhalla, J. Cafazzo, and A. Mariakakis, "On the production and measurement of cardiac sounds in the ear canal," in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2024, pp. 685–690.

[26] Thinklabs, "Thinklabs one - digital stethoscope." [Online]. Available: https://store.thinklabs.com/products/thinklabs-one-digital-stethoscope

[27] K. Christofferson, S. Bhalla, J. Cafazzo, and A. Mariakakis, "On the production and measurement of cardiac sounds in the ear canal," in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 685–690. [Online]. Available: https://doi.org/10.1145/3675094.3680526

[28] H. PA, D. G. F, B. C, N. E, L. B, G. MC, P. D, K. A, and K. N, "It'is database for thermal and electromagnetic parameters of biological tissues," 2 2022. [Online]. Available: itis.swiss/database

[29] E. Bradley and J. Sacerio, "The velocity of ultrasound in human blood under varying physiologic parameters," *Journal of Surgical Research*, vol. 12, no. 4, pp. 290–297, 1972. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022480472900248

[30] P. K. Bjarke Gårdbæk, "On the origin of cardiovascular sounds recorded from the ear," 06 2024.

[31] I. Rezek and S. Roberts, "Envelope extraction via complex homomorphic filtering," 07 1998.

[32] S. Messer, J. Agzarian, and D. Abbott, "Optimum wavelet denoising for phonocardiograms," *Microelectronics Journal*, vol. 32, pp. 931–941, 12 2001.

[33] D. Kumar, P. Carvalho, M. Antunes, R. P. Paiva, and J. Henriques, "Noise detection during heart sound recording using periodicity signatures," *Physiological Measurement*, vol. 32, no. 5, pp. 599–618, 2011.

[34] C. N. Gupta, R. Palaniappan, S. Swaminathan, and S. M. Krishnan, "Neural network classification of homomorphic segmented heart sounds," *Applied Soft Computing*, vol. 7, no. 1, pp. 286–297, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494605000694