Noise Robust Detection of Fundamental Heart Sound using Parametric Mixture Gaussian and Dynamic Programming

Achuth Rao M V\textsuperscript{1}, Shailesh BG\textsuperscript{2}, Drishti Ramesh Megalmani \textsuperscript{2}, Satish S Jeevannavar\textsuperscript{2}, Prasanta Kumar Ghosh\textsuperscript{1}

Abstract—In this work, we propose an unsupervised algorithm for fundamental heart sound detection. We propose to detect the heart sound candidates using the stationary wavelet transforms and group delay. We further propose an objective function to select the candidates. The objective function has two parts. We model the energy contour of S1/S2 sound using the Gaussian mixture function (GMF). The goodness of fit for the GMF is used as the first part of the objective function. The second part of the objective function captures the consistency of the heart sounds’ relative location. We solve the objective function efficiently using dynamic programming. We evaluate the algorithm on Michigan HeartSound and Murmur database. We also assess the algorithm’s performance using the three different additive noises—white Gaussian noise (AWGN), Student-t noise, and impulsive noise. The experiments demonstrate that the proposed method performs better than baseline in both clean and noisy conditions. We found that the proposed method is robust in the case of AWGN noise and student-t distribution noise. But its performance reduces in case of impulsive noise.

I. INTRODUCTION

Automatic detection of fundamental heart sounds from the phonocardiogram (PCG) recording is an essential step in acoustic cardiac signal analysis [1]. The detection of fundamental heart sound can assist in visualizing the heart sounds [2] and identify systolic or diastolic regions of a PCG, which helps in the diagnosis of pathological murmurs, clicks, and other heart sounds. The first heart sound (S1) occurs after the R-peak (ventricular depolarisation) in the electrocardiogram. The second heart sound occurs between the systole and diastole regions.

There are various methods in the literature that address the problem of S1/S2 detection. Some of the early methods are threshold-based, which uses a threshold on various features of the PCG signal to detect S1/S2. Some of the examples of features are Shannon energy [1], squared energy [3], frequency domain envelopograms [4], Hilbert envelope [5] and wavelet transforms [6]. However, the threshold used in threshold-based methods depends on the quality of the recordings.

To address the limitation of the threshold-based methods, there are various approaches, that use a multi-dimensional feature extraction from the PCG segments followed by different classifiers to detect the S1/S2 sounds. These features include time domain features [7] or frequency domain features [8]. There are various classifiers used including Neural Networks [9], [10], decision tree [11], support vector machines (SVM) [12], hidden Markov model (HMM) classifier [13], [14] and deep neural networks [15]. The performance of a method depends on the robustness of the features used for the classification.

The main limitation of the methods discussed above is that no temporal information is used for the S1/S2 prediction. In order to capture temporal information, Chend et al. used an ergodic HMM with Gaussian mixture model (GMM) distribution as emission density [16] to find the S1/S2 sounds. Typically, three major limitations affect the performance; of S1/S2 detection: 1) the duration distribution of S1/S2 may not match the test recording, 2) the features used may not follow the emission distribution assumption, 3) the training emission distribution may not match the test emission distribution. There are methods in the literature that handles the first and second limitations. Schmit et al. modified the hidden semi-Markov model (HSM) to get a better duration model of HMM to improve the performance [17]. An adaptive duration model methodology was proposed in [18] and [19]. The SVM and logistic regression function are also used for emission probability estimation. The HSM with logistic regression function based emission probability estimation was used as the baseline algorithm in 2016 PhysioNet Challenge. The main disadvantage of these methods is that the feature distribution may not follow the assumed distribution.

There are various methods that learn the features during training. The convolution neural network has been used to directly predict the emission probability to improve the performance [20]. One of the recent papers uses Bi-directional Long Short Term Memory Networks with attention to detect the heart sounds to improve the performance [21]. However, it has already been shown that the deep learning methods fail when the test recording conditions are different from the training conditions [22].

In this work, we propose an unsupervised method for heart sound detection. First, we detect the candidate of S1/S2 using the multi-scale product and group delay based method. We propose an objective function to select the optimal S1/S2 locations from the detected candidates. The objective function has two parts. The first part models the short term energy of the heart sound as a mixture of Gaussian functions and uses the goodness of fit of the energy contour to score the locations of S1/S2. The second parts uses a scoring function that accounts for the relative positions consecutive S1 sounds, S2 sounds, and S1/S2 locations. The maximization of this objective functions is efficiently
solved using the dynamic programming to estimate the S1/S2 locations. We evaluate the algorithm on Michigan Heart Sound and Murmur database. We also systematically evaluate the algorithm’s robustness using three different additive noise types: white Gaussian noise, Student-t noise, and impulse noise. Few algorithms in the literature study the robustness of the method in various synthetic noise conditions [10]. The synthetic noise allows us to control various parameters to understand their effect on the performance. The experiments demonstrate that the proposed method performs better than baseline in both clean and noisy conditions. We found that the proposed method is robust in cases of AWGN noise and student-t distribution noise. But its performance reduces in case of impulsive noise.

II. PROPOSED FUNDAMENTAL HEART SOUND DETECTION

The goal of heart sound detection is to find the locations of S1/S2 sounds given the heart sound signal \( s[n] \), sampled at the sampling rate \( f_s \). Fig. 1 shows the block diagram of the proposed heart sound detection. It involves two main parts. We discuss the details of each part in the next subsections.

A. Candidate detection

The main feature of S1/S2 sounds is that they are impulsive in nature, but it can be challenging to find them in low signal-to-noise ratio (SNR) conditions. Hence, we propose to estimate the candidates of S1/S2 using the multiscale product of the stationary wavelet transform. The multi-scale product has been shown to perform well in the detection of impulse like signal in low SNR conditions [23]. First, we decompose the \( s[n] \) into wavelet components as follows:

\[
d_j(n) = \sum_k g_j(k)a_{j-1}(n-k), a_{j}(n) = \sum_k h_j(k)a_{j-1}(n-k)
\]

where \( g(k) \) and \( h(k) \) are the details and approximation filters[23]. The max value of \( j \) is given by \( \log_2(N) \), where \( N \) is the number of samples in the signal. \( a_0(n) = s[n] \).

We further half wave rectify the \( p[n] \) to get the \( p^H[n] \). The resulting \( p^H[n] \) is a sparse impulse like signal as shown in Fig. 2. Its clear from the figure that the signature of S1/S2 is enhanced and less sensitive to noise/murmurs in the heart sound (Fig 2(b) and (c)). The zero crossings (negative value to positive value) in group delay computed from \( p^H[n] \) are used as the candidates of S1 or S2. The set of detected candidates is indicated by \( I \), which are shown using the black diamond in second row of Fig. 2. It is clear that the candidate detection process is robust to AWGN. It is also clear that the other heart sound, i.e., S3/S4, or impulsive noise could be detected as candidates for S1/S2. We further propose a selection technique from these candidates to filter S1/S2 locations.

B. Candidate selection

Given the candidate set \( I \), the goal of candidate selection is to find the locations of S1 (\( \Psi \)) and the locations of S2 (\( \zeta \)) by maximizing the following cost function.

\[
\Psi^*, \zeta^* = \arg\max_{\Psi, \zeta} \sum_{i=2}^{N} C(\Psi_i, \zeta_i, \Psi_{i-1}, \zeta_{i-1}) \tag{3}
\]

such that \( \Psi_i > \zeta_i > \Psi_{i-1}, 2 \leq i \leq N \)

where \( C(\Psi_i, \zeta_i, \Psi_{i-1}, \zeta_{i-1}) = \log(\mathcal{E}(\Psi_i, \zeta_i)) + \log(\mathcal{T}(\Psi_i, \zeta_i, \Psi_{i-1}, \zeta_{i-1})) \). The cost function consists of two parts. The first part is related to the goodness of fit based cost given the locations of S1 and S2. The second
part of the cost function captures the consistency in the locations of S1/S2 relative to the heart rate. We explain these two parts of the cost function in detail in the next subsections.

1) **Goodness of fit (GoF)**: We divide the \( s[n] \) into frames using a window length of \( N_w \) and the shift \( N_s \). We compute short-time energy as follows:

\[
e[n] = \sum_{m=n \times N_s + 1}^{n \times N_s + N_w} s[n]^2
\]

The short-time energy contours computed for the three examples are shown in Fig. 2. It is clear from the figure that the energy is higher in case of the sounds S1 and S2.

It is clear from the figure that the energy contour of both S1 and S2 can be jointly modeled using Gaussian mixture model (GMM) with the S1 peak at \( \mu_1 \) and S2 at \( \mu_2 \) as follows:

\[
f(x_n; \mu_1, \mu_2) = a + b \times x_n + c \times e^{-\frac{(x_n-\mu)^2}{\alpha}} + g \times e^{-\frac{(x_n-\mu_2)^2}{\eta}}
\]

The affine term \( a + b \times x \) is added to account shift in energy contour because of the murmurs. Given the energy contour and the locations of S1 and S2, we compute the minimum error.

\[
E(\mu_1, \mu_2) = \min_{a,b,c,d,g} \sum_{n \in R} (f(x_n; \mu_1, \mu_2) - e[n])^2
\]

where \( R \in [\mu_1 - l, \mu_2 + l] \). \( R \) is the range of summation that considers \( l \) samples before the given S1 location and \( b \) samples after the S2 location. We solve the optimization in eq 5 using gradient descent.

2) **S1/S2 location consistency**: The goodness of fit alone is not sufficient to differentiate the S1/S2 sounds. Hence, we define some priors on the distance between consecutive S1, S2, S1 to S2 and S2 to S1. Let the consecutive locations of S1 be \( \Psi_i \) and \( \Psi_{i-1} \) and consecutive locations of S2 be \( \zeta_i \) and \( \zeta_{i-1} \). Suppose \( H \) indicates the average distance between consecutive S1. The consistency cost of the S1/S2 locations is given by

\[
T_{S1}(\Psi_{i-1}, \Psi_i) = e^{-\alpha(\Psi_i - \Psi_{i-1} - H)^2}
\]

\[
T_{S2}(\zeta_{i-1}, \zeta_i) = e^{-\eta(\zeta_i - \zeta_{i-1} - H)^2}
\]

The distance between consecutive S1 or S2 locations may not be sufficient to differentiate the S1/S2 sounds. Hence, we further impose the location consistency constraint on the distance between S1 to S2 and S2-S1. The costs are defined as below

\[
T_{S1S2}(\Psi_i, \zeta_i) = e^{-\theta(\Psi_i - \zeta_i - \gamma)^2}
\]

\[
T_{S2S1}(\Psi_{i+1}, \zeta_i) = e^{-\eta(\Psi_{i+1} - \zeta_i - (1-\gamma)H)^2}
\]

The \( \alpha, \eta \) and \( \gamma \) controls the cycle to cycle variability in the location of consecutive S1 or S2. The distance from S1 to

\[\text{Please note that the objective function is not convex.}\]

The following S2 is typically less than \( \frac{H}{2} \) and the distance from S2 to the following S1 is typically higher than the \( \frac{H}{2} \). Hence the typical value of \( \gamma < 0.5 \). This cost can help in differentiating the S1 sound from S2 sound.

The total location consistency loss is defined as the log sum of four losses in eq 6-9.

3) **Optimization**: The maximization of eq. 3 is a combinatorial problem and has high time complexity. Hence, we propose to solve it using dynamic programming to reduce the complexity.

### III. Experiments and Results

#### A. Data sets and baseline schemes

We evaluate the heart sound detection algorithm using the Michigan Heart Sound and Murmur database (UMHS) [24]. There are 23 PCG signals in this corpus, which include four healthy subjects ad 19 pathological cases. The sampling frequency of the signal is 44.1kHz. The boundaries of S1/S2 were marked using audacity by two annotators and verified by an expert cardiologist.

We use an HMM with a logistic density-based segmentation method as a baseline, for which we use a public implementation given by the authors [25]. The HMM is a supervised method. We use a leave one out cross-validation for HMM to get the S1/S2 predictions for all recordings.

#### B. Experimental setup

We resample the recording to 2kHz as there is very little spectral content above 1kHz. We estimate the heart rate(H) using autocorrelation method. We compute the energy contour using a window size of 50ms with a shift of 10ms. We low pass filter the energy contour using a cutoff frequency of 100Hz. We set \( \gamma = 0.25 \) to have a prior position of the S2 to be in one-fourth of the cardiac period. We set the \( \alpha \) to 1, \( \theta \) and \( \eta \) to 0.1. We initialize the variable a and b to zero, c and g to 0.5, d and f to 0.1 and solve eq. 5 using gradient descent with a learning rate of \( 10^{-3} \) for 100 iterations. We indicate the proposed method by PMGDP.

We also evaluate the model in three additive noise conditions – additive white Gaussian noise (AWGN), Student-t distribution, and impulse noise. The AWGN noise added to the clean signal at an SNR level of 30dB, 10dB, 0dB, and -5dB. A typical distribution of noise observed in heart sound is high tailed in nature; it’s mainly due to the stethoscope diaphragm rubbing with the cloth or the skin. To do a

---

![Fig. 3. Comparison of sensitivity, specificity and F-score for the PMGDP and HMM for S1/S2 sounds](image-url)
systematic study of the robustness of the methods for different tail probability, we add the Student-t noise at different degrees of freedom (gamma) in the range 3,2.5,2,1.5. The extreme case of heavy-tailed noise is the impulse noise. The impulse noise of duration $D$ is added ten times to the whole signal at randomly chosen locations. For each of these ten times, a signal chunk of duration $d$ is replaced by $\{\max(|s[n]|), -\max(|s[n]|)\}$. We experiment with various values of $d$, namely, 0.1, 0.2, 0.3, 0.4 seconds. We use the HMM model trained on the clean signal to evaluate the noisy test cases.

We evaluate S1/S2 detection performance using sensitivity (Se), specificity (P+), and F-score (FS) similar to [17]. We consider that the detection is True positive if there is a prediction within S1/S2 segments. If there is more than one detection inside the S1/S2 segment, it is treated as a False-positive. Any detection outside the S1/S2 segmented is treated as False positive; no detection inside an S1/S2 segment is treated as False-negative. If there no detection within the S1/S2 segment is treated as True negative.

C. Results on clean heart sound

Fig. 3 shows the comparison of sensitivity, specificity, and F-score for the PMGDP and HMM for S1/S2 sounds. It is clear from the figure that the PMGDP performs better than HMM in S1 in terms of all three metrics. In the S2 sound, the sensitivity values of both the methods are comparable, and the PMGDP method’s specificity is better than that using HMM. It could be mainly because of the lack of training data for the HMM, where it detects some of the rare events such as S3, S4 sounds, or systolic click as S2. We found that the PMGDP can eliminate the S3/S4 events mainly because of the location consistency part in the cost function. The F-Score is very poor for PMGDP when the S2 is missing in the test recording, where it detects S1 or any other noises as S2.

D. Results on noisy recordings

Fig. 5 shows the F-score for different noises and the noise parameters. The Fig. 5(AWGN) shows the F-score using HMM and PMGDP for S1S2 sounds at different SNR levels. It is clear from the figure that the F-score for HMM is poor mainly because the HMM is failing to generalize to an unseen noise. The F-score using PMGDP for both S1/S2 decreases as SNR decrease. The Fig. 5(student-t) shows the F-score of HMM and PMGDP for S1/S2 sounds at different gamma value. It is clear from the figure that the performance of the PMGDP decreases with the degree of freedom. The lower degrees of freedom causes more outliers in the signal, making it challenging for the algorithm to detect the S1/S2 sounds. The Fig. 5(impulse-noise) shows the F-score of HMM and PMGDP for S1S2 sounds at the different duration of impulse noise. But the performance of PMGDP drastically
reduced at the small duration of the noise (0.1s) because the GoF fit biases the predictions towards the noisy regions. We also found that the HMM performance is poor even in those cases.

Fig. 4 shows an illustrative example of PMGD based S1/S2 detection for clean and noisy sounds. In the case of clean sounds, the detection of S1/S2 is accurate in most cases. In case of (a), the detection is good even though the strength of S1/S2 is comparable. In case of (b), the S1 is detected accurately even though there is S4 sound. In case of (d), the S2 is detected accurately even with the split S2 signal. In case of (c), the prediction of S2 is wrong, and S1 is missed. It is mainly because the candidate selection is failing to detect the S2 as a candidate. In the case of (e), the PMGD can detect the S1/S2 with the 0dB noise. In the case of (g&h), the detected locations are accurate even though there is outlier noise. If the outlier noise magnitude is too high, the GoF part dominates the objective function resulting in the false detection.

IV. Conclusion

In this work, we proposed an unsupervised two-step method for fundamental heart sound detection. The first step involves the detection of heart sound candidates using wavelet transform and group delay. In the second step, the objective function is optimized using dynamic programming. The objective functions consider the cost of fitting Gaussian mixture function to the energy contour and relative location of S1/S2. We evaluated the method using Michigan Heart-Sound and Murmur database. The experiments demonstrated that the proposed method achieves a better F-score compared to a supervised HMM-based method in clean as well as in noisy conditions. As part of future work, we plan to include the costs based on the spectral content of S1/S2 for the robust detection in the case of impulsive noise.

REFERENCES