Logistic Regression-HSMM-Based Heart Sound Segmentation

David B. Springer*, *Student Member, IEEE*, Lionel Tarassenko, *Senior Member, IEEE*, and Gari D. Clifford, *Senior Member, IEEE*

***Abstract*—The identification of the exact positions of the first and second heart sounds within a phonocardiogram (PCG), or heart sound segmentation, is an essential step in the automatic analysis of heart sound recordings, allowing for the classification of pathological events. While threshold-based segmentation methods have shown modest success, probabilistic models, such as hidden Markov models, have recently been shown to surpass the capabilities of previous methods. Segmentation performance is further improved when *a priori* information about the expected duration of the states is incorporated into the model, such as in a hidden semi-Markov model (HSMM). This paper addresses the problem of the accurate segmentation of the first and second heart sound within noisy real-world PCG recordings using an HSMM, extended with the use of logistic regression for emission probability estimation. In addition, we implement a modified Viterbi algorithm for decoding the most likely sequence of states, and evaluated this method on a large dataset of 10 172 s of PCG recorded from 112 patients (including 12 181 first and 11 627 second heart sounds). The proposed method achieved an average F_1 score of $95.63 \pm 0.85\%$, while the current state of the art achieved $86.28 \pm 1.55\%$ when evaluated on unseen test recordings. The greater discrimination between states afforded using logistic regression as opposed to the previous Gaussian distribution-based emission probability estimation as well as the use of an extended Viterbi algorithm allows this method to significantly outperform the current state-of-the-art method based on a two-sided paired t-test.**

***Index Terms*—Heart sound segmentation, hidden Markov models (HMMs), logistic regression (LR), phonocardiography (PCG).**

I. Introduction

THE segmentation of the fundamental heart sounds (FHSs) is an essential step in the automatic analysis of the phonocardiogram (PCG). The accurate localization of the FHSs is a prerequisite for the identification of the systolic or diastolic regions of a PCG, allowing the subsequent classification of pathological murmurs in these regions [1]. The FHSs refer to the first heart sound, S_1, and the second heart sound, S_2, originating at the beginning of mechanical systole and diastole, respectively [2]. S_1 occurs immediately after the R-peak (ventricular depolarization) of the electrocardiogram (ECG), while S_2 occurs at approximately at the end-T-wave of the ECG (the end of ventricular depolarization) [3], as shown in Fig. 1.

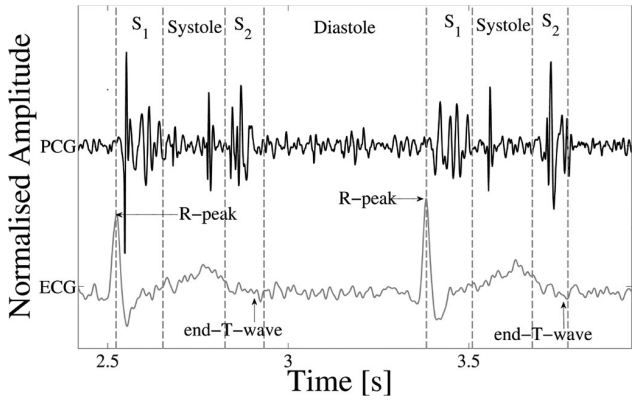

Fig. 1. Example of an ECG-labeled PCG, with the ECG, PCG, and four states of the heart cycle (S_1, systole, S_2, and diastole) shown. The R-peak and end-T-wave are labeled as references for defining the approximate positions of S_1 and S_2, respectively. Midsystolic clicks, typical of mitral valve prolapse, can be seen.

While the segmentation of heart sounds is relatively simple in noise-free recordings, it becomes a difficult task when the recordings are corrupted by in-band noise. Common noise sources include endogenous or ambient speech, motion artefacts, and physiological sounds, such as intestinal and breathing sounds. Other physiological sounds of interest, such as murmurs, clicks, splitting of the FHSs, and additional S_3 and S_4 sounds, can also complicate the identification of the FHSs.

This paper addresses the problem of accurate segmentation of the FHSs in noisy real-world recordings from healthy and pathological patients without the use of a reference signal, such as an ECG. The principal contributions of this paper are: an exploration of features for heart sound segmentation, including a robust selection of the wavelet family and decomposition level when using the discrete wavelet transform (DWT); an enhanced hidden Markov model (HMM), which includes duration dependencies and logistic regression (LR)-based emission probabilities; and the implementation of an extension to the Viterbi algorithm for use with hidden semi-Markov models (HSMMs). The proposed segmentation approach is rigorously evaluated on one of the largest published datasets, and compared to the current state-of-the-art segmentation algorithm.

II. Background

Table I summarizes the relevant background literature. The table presents the size of the datasets used, numerical results

Manuscript received May 21, 2015; revised July 21, 2015; accepted August 19, 2015. Date of publication September 1, 2015; date of current version March 17, 2016. The work of D. B. Springer was supported by the Rhodes Trust. *Asterisk indicates corresponding author.*

*D. B. Springer is with the Institute of Biomedical Engineering, University of Oxford, Oxford OX3 7DQ, U.K. (e-mail: david.springer@eng.ox.ac.uk).

L. Tarassenko is with the University of Oxford.

G. D. Clifford is with the University of Oxford, Emory University, and also the Georgia Institute of Technology..

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2015.2475278

TABLE I
SUMMARY OF PREVIOUS WORK

Authors	Dataset	Reported Metrics and Results	Notes
[4]	37 recordings (515 cycles) from children with murmurs (14 being pathological)	93.0% Ac	Unsupervised. Optimized on entire dataset
[5]	77 (1165 cycles) recordings from children with both pathological and physiological murmurs	94.6% Ac	Unsupervised. Optimized on entire dataset
[6]	55 recordings (7530 cycles), 51 with valve replacements	97.95% Se, 98.2% Sp	Unsupervised. Optimized on entire dataset
[7]	71 recordings (357 cycles), nine different pathologies	97.47% Ac	No split between train, test sets
[8]	166 clean heart cycles from normal and pathological patients	84.0% Ac	Unsupervised. No stated segmentation tolerance
[9]	41 recordings (340 cycles). Mix of normal (32%), systolic (36%) and diastolic murmurs (32%)	90.29% Ac	Unsupervised
[10]	27 recordings of 30 s (997 cycles) from healthy subjects	92.1% Se, 88.4% P_+	Unsupervised
[11]	30 clean recordings (20 s) from healthy subjects	96.2% Ac	No split between train, test sets
[12]	120 recordings from children, 80 with congenital heart disease (totaling 1200 s, 823 cycles in test set)	93.6% Ac on test set	50% train-test split
[13]	Nine recordings (less than 5 s). 55% pathological	99.0% Ac on whole cycle detection	No split between train, test sets
[14]	9426.8 s of recordings, normal (22.2%) and various pathologies (ASD, PDA, VSD, and RHD)	S_1: 98.53% Ac, S_2 : 98.31% Ac, Cycles: 97.37% Ac	Unsupervised. No stated segmentation tolerance.
[15]	80 recordings from an unknown number of patients of 6–12 s (40 healthy, 40 pathological recordings)	96% and 97% Se, 95% and 95% P_+ (healthy and pathological)	Unsupervised algorithm. No stated segmentation tolerance
[16]	26 clean recordings (565 cycles), 3 healthy subjects, and 23 with various pathologies	94.9% and 95.9% Ac (S_1 and S_2)	No split between train, test sets and no stated segmentation tolerance
[17]	50 2-min healthy and pathological recordings	99.0% Se and 98.6% P_+	No split between train, test sets. Results reported on 20% of dataset
[18]	64 teaching quality recordings of less than 10 s (701 cycles). Various pathologies	93.06% Ac, 99.43% Se. 93.56% P_+	No split between train, test sets. Results reported on portion of dataset. No stated segmentation tolerance.
[19]	52 recordings (14 controls, 38 with murmurs), 43 in test set (2602 cycles)	83.05 $\pm$ 15.14% Ac94.56 $\pm$ 6.58 $G-measure$	Ac denoted for correctly segmented cycles. $G-measure$ is geometric mean of Se and P_+
[20]	80 patients, 8 pathological. Recordings of 20 s from four auscultation sites (10045 S_1, 9818 S_2 sounds	S_1: 94.6% Se and 97.7%P_+ S_2 : 95.2% Se and 96.1% P_+	No split between train, test sets
[21]	46 clean recordings from eight patients (2286 s). No pathologies mentioned	97.6% Ac	Ac computed from average of eightfold cross validation
[22]	17 patients, 44 recordings (30–60 s). No pathologies mentioned	S_1: 98.6% Se and 96.9% P_+ S_2 : 98.3% Se and 96.5% P_+	Results computed from average of fourfold cross validation
[23]	113 recordings of 8 s, 8% with coronary artery disease	98.8% Se, 98.6% P_+ on test set	73 test, 40 training recordings

Ac denotes accuracy, Se sensitivity, Sp specificity, and P_+ positive predictivity.

of the studies, and important notes, such as whether any independent test set was used to evaluate the presented algorithm.

Many methods of heart sound segmentation use an amplitude threshold after various transformations of the PCG signal. Numerous researchers have applied this approach with various features [4]–[9]. Chen *et al.* [10] used a portion of the dataset used in this paper, and adapted ECG analysis methods based on the use of a threshold and k-means clustering to identify the heart sounds. Others have used neural networks to segment FHSs, using Morlet wavelet decomposition or frequency band and periodicity features [11], [12]. Yan *et al.* [13] derived a characteristic moment waveform, based on the variance of the PCG over differing time scales, and showed promising results on a small dataset. Sun *et al.* [14] used the same envelope as Yan *et al.* [13], but then from this derived a modified Hilbert transform-based envelope. Moukadem *et al.* [15] found the Shannon envelope after applying the S-transform to locate the FHSs, while Tang *et al.* [16] employed dynamic clustering after atom decomposition, but both did not state a localization tolerance. Naseri and Homaeinezhad [17] employed a combined frequency–energy feature, while Varghees and Ramachandran [18] employed the phase from the analytical signal after finding the Shannon entropy, but both optimized their approach on their entire dataset and reported results on a small portion of this. Papdaniil *et al.* [19] outperformed many of these methods using empirical mode decomposition and kurtosis features to select non-Gaussian intrinsic mode functions (IMFs), and detected the start and end positions of heart sounds within the selected IMFs.

The introduction of probabilistic models for heart sound segmentation led to improved accuracy. Gamero and Watrous [20] applied two separate HMMs, trained using mel-frequency cepstral coefficients, to a large dataset of mostly healthy patients with notable success. Ricke *et al.* [21] used embedded HMMs within each state, along with mel-frequency cepstral coefficients, Shannon energy, and regression coefficients on a small dataset with success. Gill *et al.* [22] were the first researchers to consider timing durations within HMMs for heart sound segmentation, incorporating the time to preceding and following amplitude peaks in the homomorphic envelope as input features into the HMM, which yielded high segmentation accuracy on a small dataset. Schmidt *et al.* [23] were the first researchers to explicitly model the expected duration of heart sounds within

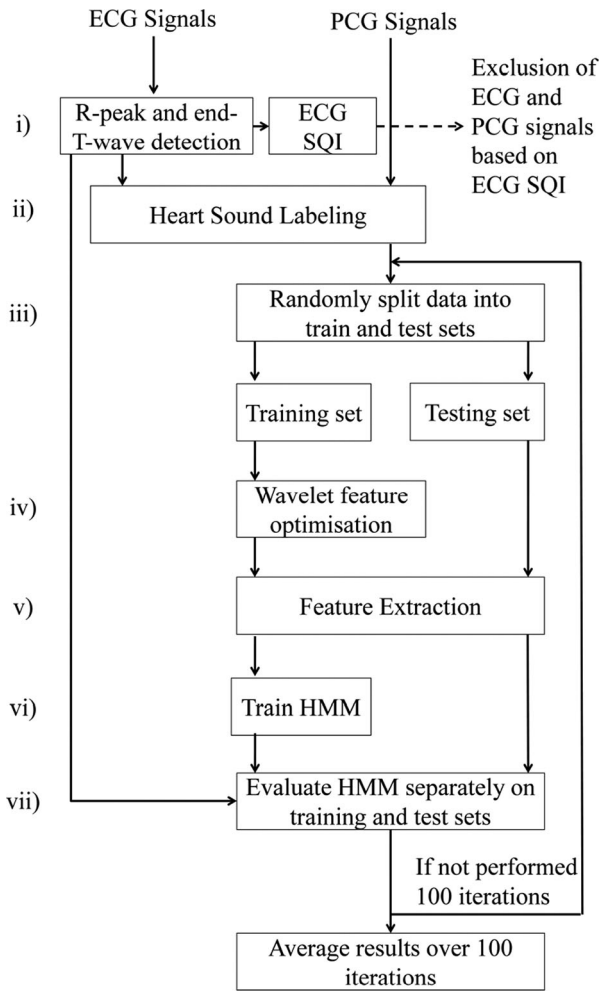

Fig. 2. Block diagram of the steps employed in the evaluation of the segmentation algorithms.

the HMM using an HSMM. These researchers performed a rigorous assessment of this method on a noisy dataset and demonstrated this method to be the current state-of-the-art. However, limitations of this paper include a small dataset of only 113, 8-s recordings, and the investigation of only two input features.

This paper builds on the work of these researchers by using a substantially larger dataset, the assessment of additional input features, an extension of the Viterbi algorithm, and more discriminative emission probability derivation using LR within the HSMM.

III. Methods

The flow diagram in Fig. 2 illustrates the different steps used in the evaluation of the heart sound segmentation methods.

A. Study Database

The study database consisted of 405 synchronous 30–40-s PCG and ECG recordings from 123 deidentified adult patients (roughly three recordings per patient) attending the Massachusetts General Hospital for cardiac screening or in-home recordings of people suffering from mitral valve prolapse (MVP). Recordings were made using a Meditron (NY, USA) electronic stethoscope, saved in uncompressed wave format at 44.1 kHz at 16-bit resolution [24]. Many of these PCG recordings were corrupted by various sources of noise, including stethoscope motion, breathing, intestinal sounds, and talking. Comparing the size of this database to others in Table I, one can see that this dataset represents one of the largest published in terms of patient numbers and recording length.

Out of the 123 patients, 38 were normal controls, 37 had murmurs relating to MVP, 36 had benign murmurs, 5 had aortic disease, and 7 had other miscellaneous conditions (tricuspid regurgitation, endocarditis, asymmetric septal hypertrophy). Recordings were made from the parasternal, apical, aortic, and pulmonic auscultation positions, with patients sitting forward, squatting or in the supine position. All diagnoses were verified by a single clinical expert through echocardiographic examination.

B. Data Exclusion Based on ECG Signal Quality

This section refers to step *i)* in Fig. 2. The gold-standard reference positions for the FHSs in the PCG were derived from the synchronous ECG recordings. The R-peak and the end-T-wave in the ECG correspond to the positions of the S_1 and S_2 sounds in the PCG, respectively [3] (see Fig. 1) and were the reference positions for the PCG labeling in Section III-D and evaluation of the segmentation algorithms in Section IV (as shown in steps *ii)* and *vii)* in Fig. 2). Therefore, it was essential to ensure the correct detection of the R-peak and the end-T-wave in the ECG. The correct positions of these markers in the ECG were found by comparing the agreement between four R-peak and four end-T-wave detectors. Intuitively, the presence of noise and artefacts within the ECG will lead to disagreement between semi-independent detectors for both the R-peak and end-T-wave. Therefore, in order to ensure high-quality periods of the ECG signal and accurate localization of the R-peak and end-T-wave, the agreement between the R-peak and end-T-wave detectors was assessed in order to derive an ECG signal quality index (SQI).

The four R-peak detectors employed in this study were "gQRS" (available on Physionet [25]), "jQRS" (previously used in two studies [26], [27]), an algorithm based on parabolic fitting [28], and a wavelet-based ECG delineator [29]. The four end-T-wave detectors employed in this study were "ecgpuwave" [30] (available on Physionet [25]), a method based on maximizing the area in a sliding window between successive R-peaks [31], the same wavelet-based delineator mentioned previously [29], and a method developed by Vazquez-Seisdedos *et al.* [32] based on maximizing the area of a trapezium fixed at points within the ECG.

The performance of R-peak detectors is usually assessed by beat-to-beat comparisons between the detected beats and the reference beats. The standard adult tolerance window for candidate R-peaks is 150 ms [33]. However, this is longer than the expected duration of a heart sound [23] and was believed to be too wide a tolerance in this case. In order to get a more robust estimate of the correct location of peaks, in this study, if R-peak

detectors agreed to within 100 ms, they were said to have agreed on the position of the R-peak. The same tolerance was followed for the end-T-wave positions.

The process for R-peak detection was:

1) Bandpass filter the ECG signal between 0.7 and 50 Hz using zero-phase forward–backward second-order Butterworth filters to ensure no baseline wander and the exclusion of high-frequency noise.
2) Measure the agreement between all pairs of R-peak detectors, where agreement is measured as per the "bxb" algorithm, available from Physionet [25], which evaluates the agreement between two sets of annotations within a specified tolerance. The agreement was measured using the F_1 score (see Section III-I3). Then, select the three R-peak detectors that had the highest product of their F_1 scores, thereby excluding the R-peak detector that had the lowest overall agreement with the other detectors over the entire record. This allows the exclusion of a detector that may perform poorly on a particular record while keeping three independent R-peak detectors.
3) Over four second windows, the ECG SQI was labeled as the F_1 score of agreement between the chosen three detectors. The choice of a 4-s window was based on the need for at least two heart cycles, even in the case of low heart rates. Two heart cycles are needed, as many of the end-T-wave detectors employed subsequently rely on the detection of at least two R-peaks. This is equivalent to methods employed by other authors, with a smaller window size [34], [35]. The window was then shifted by 1 s, leading to a second-by-second SQI.
4) In windows of 100% F_1 score, or complete agreement between the three peak detectors within the 100-ms tolerance, the adjudicated position of the R-peaks were defined to be the R-peak position from the three detectors with the maximum absolute value. Windows of less than 100% F_1 score were not used for further analysis, as they were deemed to have untrustworthy annotations.

The process for end-T-wave detection was:

1) In windows of 100% F_1 score (based on the R-peak detections), find the end-T-wave positions from the four detectors.
2) Exclude the annotation furthest from the median of these four annotations, allowing the exclusion of an outlier annotation. If only three end-T-wave detections were present, due to a missed detection, no exclusion was performed. If fewer than three annotations were present, a missed end-T-wave was marked and the ECG SQI was set to zero.
3) If the remaining three annotations were all within the 100-ms tolerance of each other, the adjudicated end-T-wave position was marked as the median position of the three annotations. Otherwise, the ECG SQI was set to zero to indicate an untrustworthy portion of the signal.

ECG signals, and the corresponding PCG signals, of at least two heart cycles of continuous 100% SQI were kept for further analysis. Periods of ECG without 100% SQI, based on the lack of agreement between R-peak and end-T-wave annotations, were excluded as their annotations were possibly erroneous.

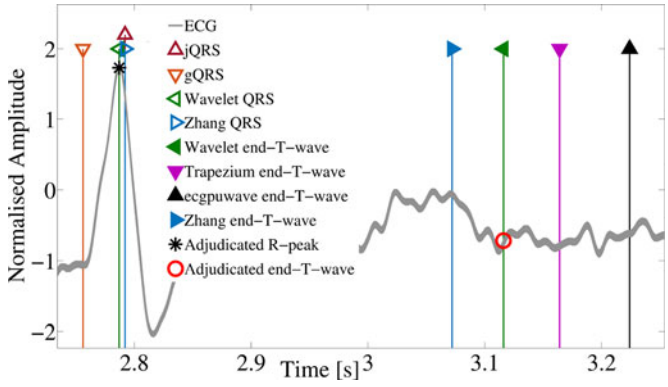

Fig. 3. Example of the accurate R-peak and end-T-wave positions derived in a noisy ECG signal based on the agreement between R-peak and end-T-wave annotations. The positions of the four R-peak detections and four end-T-wave detections are shown. The algorithm adjudicated R-peak and end-T-wave, based on the four detections, is shown.

Two cycles of continuous high-quality ECG were selected as this would give the HSMMs employed in this study sufficient data to accurately assign state labels. This is because such segments of data would consist of at least one complete systolic and diastolic period, the durations of which aid in the differentiation of the S_1 and S_2 sounds, as systole is expected to be shorter than diastole.

The remaining data consisted of 10 171.85 s of PCG, including 12 181 R-peaks and 11 627 end-T-waves (corresponding to the same number of S_1 and S_2 sounds), from 112 patients. An example of the R-peak and end-T-wave detection in the ECG is shown in Fig. 3. It is important to note that the ECG signal and the derived R-peak and end-T-wave locations were only used for heart sound labeling and evaluation of the segmentation models. They were not used in the heart sound segmentation process itself. Furthermore, ECG signal quality has little to no impact on the PCG signal quality, except in the case of motion artifact, which is limited in such a dataset of recordings performed on adults. Therefore, selection of high-quality ECG does not bias this study toward the selection of high-quality PCG recordings.

C. Heart Sound Duration Distributions

A key component of the HSMM models used in this study is an estimate of the probability density function of the time expected to remain in each state. In the case of heart sound analysis when using a four-state HMM, this is the duration in each of the four major components of a heart cycle. These are: 1) the S_1 sound, 2) the systolic period between S_1 and S_2 sound, 3) the S_2 sound, and 4) the diastolic period between S_2 and S_1 (see Fig. 1). The duration of each of these components was modeled as described by Schmidt *et al.* [23], who modeled the duration of each state as a Gaussian distribution modeled on their own annotated dataset. These distributions are heart rate dependent. Therefore, the reference heart rate for each PCG was

calculated using the R-peak locations computed in Section III-B. This section is not illustrated in Fig. 2, but the durations are intrinsic to steps *ii)* and *vii)* of the flow diagram.

D. ECG-Derived Heart Sound Labeling

In order to train the segmentation algorithms on the PCG data, it was necessary to label the S_1 and S_2 sounds within the PCG recordings using an external reference signal. The references for the positions of the S_1 and the S_2 sounds within the PCGs were the R-peak and end-T-wave positions that were computed from the synchronous ECG signals in Section III-B. This is shown as step *ii)* in Fig. 2.

As the start of S_1 coincides with the R-peak of the ECG [3], the period from each R-peak to the mean S_1 duration ($\overline{S_1}$, found from [23] in Section III-C) from each R-peak was labeled as a S_1 sound.

The S_2 sound occurs at approximately the end-T-wave in the ECG [3]. Therefore, labeling only based on the end-T-wave position would be erroneous. However, the amplitude of the S_2 sound would reach a maximum in the vicinity of the end-T-wave. Therefore, the center of the S_2 sound was found by searching for the maximum peak in the Hilbert envelope (see Section III-H) of the PCG signal within a specified search window around the end-T-wave. This search window was set to the position of the end-T-wave, plus and minus the longest expected duration of S_2 ($\overline{S_2} \pm \sigma_{S_2}$). The position of the maximum value of the Hilbert envelope within this search window was marked as center of the S_2 sound. An interval equal to the length of the expected S_2 sound, centered on this maximum position, was labeled as the S_2 sound. The period between the S_1 and S_2 sounds was labeled as systole, while those periods between S_2 and S_1 were labeled as diastole. An example of a labeled PCG recording is shown in Fig. 1.

E. Hidden Semi-Markov Models

HMMs are a statistical framework used to describe sequential data. They operate by making inferences about the likelihood of being in certain discrete "hidden states," moving between those states and seeing an observation generated by each state. In this paper, the HMM was first order, with the hidden sequence consisting of the four states of the heart (S_1, systole, S_2, diastole), while the observed sequence is the PCG signal, or features computed from the PCG signal. An HMM is governed by the "Markov property," which states that the next state is only dependent on the state that is occupied at the current time step. This is valid for heart sounds, as each successive state can only be reached from one particular previous state. A HMM can be defined as

$$\lambda = (A, B, \pi) \tag{1}$$

where A is the transmission matrix, B is the emission or observation distribution, and π is the initial state distribution [36]. Let the hidden states be defined as $\xi = [\xi_1, \xi_2, \ldots, \xi_N]$, where N is the total number of states. In this paper, $N = 4$ and ξ_1 refers to S_1, ξ_2 refers to systole, ξ_3 refers to S_2, and ξ_4 refers to diastole. Let the duration of an entire sequence be T and the state at time t be q_t with the entire sequence of states being $\mathbf{Q}$ [36]. Let the observation sequence be $\mathbf{O} = \{\mathbf{O}_1, \mathbf{O}_2, \ldots \mathbf{O}_T\}$, where $\mathbf{O}_t$ is a vector of feature values at time t.

$A = \{a_{ij}\}$ defines the probability of moving from state i at time t to state j at time $t+1$. A four-state heart sound HMM is a case of a "nonergodic" HMM, where each state is only accessible from one specific to other state. For example, S_2 has to precede diastole; diastole cannot be preceded by systole or S_1.

$B = \{b_j(\mathbf{O}_t)\}$, $1 \leq j \leq N$ defines the probability of state j that generates the observation vector $\mathbf{O}_t$ at time t.

The initial state distribution $\pi = \{\pi_i\}$ defines the probability of being in state i at the first time point.

In the case of heart sound segmentation, a utility of the HMM is the computation of the optimal state sequence, given a model λ and an observation sequence $\mathbf{O}$, where optimality is defined as the most likely sequence of states. Cycling through every combination of $\mathbf{Q}$ or order to find the optimal sequence is infeasible, even for short-time sequences. Therefore, a dynamic programing method called the Viterbi algorithm is employed to solve the most likely state sequence [36].

We define $\delta_t(j)$ as the likelihood of the most probable state sequence that accounts for the first t observations and ends in state j at time t, while $\delta_1(j) = \pi_j b_j(\mathbf{O}_1)$. Incorporating the information from the previous time step, $\delta_t(j)$ can be calculated by induction using

$$\delta_t(j) = [\max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}] \cdot b_j(\mathbf{O}_t). \tag{2}$$

The argument which maximized (2), which is needed to track the optimal state sequence, is stored in the matrix $\psi_t(j)$

$$\psi_t(j) = [\arg \max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}]. \tag{3}$$

This matrix stores the most likely previous state i at time t, if in state j at time $t+1$. This allows the backtracking of the most likely sequence of states q_t^* when reaching the end of the sequence using

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \tag{4}$$

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \ldots 1. \tag{5}$$

A major limitation of the standard HMM is that it does not explicitly incorporate any information about the expected duration of each state. Without incorporating this information, the state durations are governed only by the self-transition probability a_{ii}. This results in a geometric distribution for the duration expected to remain in each state [36]. This distribution monotonically decreases, resulting in the most likely state duration always being one-time step. This is poorly suited for PCG analysis. In order to improve the duration modeling, an extra parameter is needed in the model.

Let us define the new model as $\lambda = (A, B, \pi, p)$, where $p = \{p_i(d)\}$ is the explicitly defined probability of remaining in state i for duration d (derived in Section III-C).

Then, modifying the Viterbi algorithm to include the duration densities, we find [23], [37]

$$\delta_t(j) = \max_d \left[\max_{i \neq j} \left[\delta_{t-d}(i) \cdot a_{ij} \right] \cdot p_j(d) \cdot \prod_{s=0}^{d-1} b_j(\mathbf{O}_{t-s}) \right] \tag{6}$$

with $1 \leq i, j \leq N$, $1 \leq t \leq T$, and $1 \leq d \leq d_{\max}$ where in this case $d_{\max}$, the maximum time expected to remain in any one state, is set to the duration of an entire heart cycle. This is done to ensure tractability of the algorithm.

The observation density $\prod_{s=1}^{d} b_j(\mathbf{O}_{t-s})$ is now the calculation of the probability of observing all the observations from time $t-d$ to time t in state j. It should be noted that when incorporating duration densities as in (6), the entries for the transition matrix a_{ij} in the case of a four-state heart sound model become unity for the transition from one state to another, provided it is a permissible transition (for example, S_1 to systole, or S_2 to diastole). This is due to the fact that the transition between states is now only dependent on the duration remaining in each state, and no longer on the probability of transitioning between states.

Equation (6) is maximized according to two arguments i and d, which are stored in matrices $\psi_t(j)$ and $D_t(j)$. The psuedocode for the standard backtracking procedure for the HSMM is shown in [23]. This is then called an HSMM,[1] as only the transition between different states is Markovian [37]. Since the human heart has clear upper and lower bounds on the duration of the components of the heart cycle (because of mechanical limitations), we expect that the incorporation of such information should help to improve segmentation performance.

F. Extended Viterbi Algorithm

The standard backtracking procedure for decoding the most likely sequence of states when using an HSMM, as pointed out by Schmidt *et al.* [23] and Yu [38], is limited by the fact that states are required to start and end at the start and end of the PCG signal. This is infeasible, as a PCG recording can begin and end at any stage of the heart cycle. In order to a resolve this, an extended Viterbi algorithm which extends beyond the beginning and end of the PCG is proposed, as shown in Algorithm 1.

In short, this algorithm allows the possible state durations to extend beyond the beginning and end of the PCG sequence, while only considering observations from within the PCG for emission probability estimation. This algorithm uses the equations for the "general assumption" of the forward–backward algorithm and Viterbi algorithm from Yu [38].

G. LR HSMM

A further modification to the HSMM we introduced was to incorporate LR into the model. LR-derived emission or observation probability estimates were used instead of Gaussian or Gamma distributions as employed in related work [23], [37], as the incorporation of LR into the HMM should allow for greater discrimination between states. This is similar to the use of support vector machine-based emission probabilities [39].

Algorithm 1. The extended Viterbi algorithm for use with HSMMs

$\delta_1(j) = \pi_j b_j(\mathbf{O}_1) \quad 1 \leq j \leq N$
for $t = 2 : T + d_{\max} - 1$
  for $i, j = 1 : N$
    for $d = 1 : d_{\max}$
      $start_t = t - d$
      if $start_t < 1$
        $start_t = 1$
      elseif $start_t > T - 1$
        $start_t = T - 1$
      end
      $end_t = t$
      if $end_t > T$
        $end_t = T$
      end
      $\delta_t(j) = \cdots$
  $\max_d \left[\max_{i \neq j} \left[\delta_{start_t}(i) \cdot a_{ij} \right] \cdot p_j(d) \cdot \prod_{s=start_t}^{end_t} b_j(\mathbf{O}_s) \right]$
      $D_t(j) = \cdots$
  $\arg\max_d \left[\max_{i \neq j} \left[\delta_{start_t}(i) \cdot a_{ij} \right] \cdot p_j(d) \cdot \prod_{s=start_t}^{end_t} b_j(\mathbf{O}_s) \right] \psi_t(j) = \arg\max_{1 \leq i \leq N} \left[\delta_{t-D_t(j)}(i) a_{ij} \right]$
    end
  end
end
$T* = \arg\max_t [\delta_{t=T:T+d_{\max}-1}(i)] \quad 1 \leq i \leq N$
$q^*_{T*} = \arg\max_i [\delta_{T*}(i)]$
$t = T^*$
while $t > 1$
  $d* = D_t(q^*_t)$
  $q_{t-d^*:t-1} = q^*_t$
  $q^*_{t-d^*-1} = \psi_t(q^*_t)$
  $t = t - d^*$
end

LR is a binary classification model that maps predictor variables, or features, to a binary response variable using the logistic function. The logistic function $\sigma(a)$ can be defined by $\sigma(a) = \frac{1}{1+\exp(-a)}$ [40].

Using the logistic function above, the probability of a specific class or state, given the input features or observations, can be defined by

$$P[q_t = \xi_j | \mathbf{O}_t] = \sigma(\boldsymbol{w}' \mathbf{O}_t) \tag{7}$$

where $\boldsymbol{w}$ are the weights of the model, applied to each input feature, and trained using iteratively reweighted least squares on the training data.

A one-versus-all approach was implemented, training one LR model for the observations from each state in the model. Thereafter, the probability of an observation given a state $b_j(\mathbf{O}_t|\xi_j)$, as required for the HSMM, was found using Bayes' rule

$$b_j(\mathbf{O}_t) = P[\mathbf{O}_t | q_t = \xi_j] = \frac{P[q_t = \xi_j | \mathbf{O}_t] \times P(\mathbf{O}_t)}{P(\xi_j)} \tag{8}$$

[1] Also known as an explicit duration, variable-duration, segmental or duration-dependent HMM [38].

where $P(\mathbf{O}_t)$ is found from a multivariate normal (MVN) distribution computed from the features from the entire training set and $P(\xi_j)$ is found from π, the initial state probability distribution. The inputs to the LR models are the signal-derived features described in Section III-H.

H. Feature Extraction

The segmentation algorithms used a combination of four input features (This is step *v)* in Fig. 2).

Before feature extraction, all recordings were downsampled to 1 kHz using a polyphase antialiasing filter. The majority of the frequency content of the FHS (S_1 and S_2) is below 150 Hz [41], and hence, the Nyquist–Shannon sampling criterion [42] was satisfied. The following features were then calculated.

1) Homomorphic Envelogram: The homomorphic envelogram, derived by exponentiating the low-pass-filtered natural logarithm of a signal [43], has been used by numerous researchers for extracting the envelope of PCG signals [9], [22], including the state-of-the-art segmentation algorithm [23].

2) Hilbert Envelope: Messer *et al.* [44] and Kumar *et al.* [45] calculated the envelope of the PCG signal using the Hilbert transform. The Hilbert transform extracts the analytic signal which excludes the negative-frequency components of a signal [43], and the Hilbert envelope is calculated from the absolute value of the Hilbert transform.

3) Wavelet Envelope: Wavelet analysis of heart sounds has been extensively explored. However, there is limited agreement on the optimal wavelet to use when denoising, segmenting, or classifying heart sounds. Some researchers have argued that the Morlet family is best suited for heart sound analysis [11], [46], while others have rationalized the use of the Daubechies wavelet family [5], [9], [44], [47]–[49].

In this study, the choice of wavelet was determined experimentally using the labeled heart sounds computed from Section III-D. This is shown as step *iv)* in Fig. 2.

The heart sound recordings in the training set (see Section III-I1) were decomposed using the discrete wavelet transform (DWT) with various wavelet families and decomposition levels.[2] The absolute value of the detail coefficients for each wavelet family and decomposition level were computed, in order to exclude frequency content outside of the target wavelet range and to extract a positive-valued envelope. The envelope values were summed for each state in the heart sound recordings. The wavelet family and decomposition level that yielded the highest ratio of the sum of the detail coefficients for the S_1 and S_2 sounds compared to the sum over other intervals across all recordings was selected for further use. This ratio gives an indication of how well each wavelet discriminates between the FHSs and other regions of the heart sound recordings. Therefore, a wavelet envelope computed using such a wavelet would provide the best overall discrimination between the FHSs and other sounds or noise for all heart sound recordings.

[2]These included Haar, Daubechies, symlet, Coiflet, biorthogonal, and reverse biorthogonal wavelets at decomposition levels 1–10. Morlet wavelets were not used since a DWT cannot be performed using the Morlet wavelet.

4) Power Spectral Density (PSD) Envelope: The majority of the frequency content of the S_1 and S_2 sounds is below 150 Hz with a peak at 50 Hz [41]. Based on these frequencies, the final feature was derived from the mean PSD between 40 and 60 Hz, found in overlapping windows of 0.05 s in width with 50% overlap. This resulted in an envelope of PSD values. The window size used ensures that the frequency content of the shortest expected FHS (0.05 s) is covered by an analysis window. The PSD was calculated using the short-time Fourier transform after Hamming windowing.

The feature vectors for each recording were individually normalized by subtracting their mean and dividing by their standard deviation. Following Schmidt *et al.* [23], after feature extraction and normalization, the resulting feature vectors were downsampled further to 50 Hz using a polyphase antialiasing filter in order to increase the speed of computation.

I. Model Training and Evaluation

The parameters of the HSMM model were trained using the labeled PCG sequences described in Section III-D, which were divided as follows.

1) Training and Test Data Split: In order to avoid overtraining of the model, the dataset was randomly halved into training and test sets (see step *iii)* in Fig. 2). The number of recordings for each condition (normal, murmurs relating to MVP, benign murmurs, aortic disease, and miscellaneous conditions) was split between the two sets in equal proportion ensuring stratification by patient. This ensured that no recordings from any patient were in both the training and test sets and a balanced representation of abnormalities in both training and test sets.

2) Training: The transition matrix probabilities a_{ij} and the emission probabilities B were optimized on all data in the training dataset. In the case of B, two methods were used.

In the case of the MVN emission distributions, an MVN distribution was computed for each state by finding the means and covariances of the input features for each state from the training recordings. This can be defined as $b_j(\mathbf{O}_t) \sim \mathcal{N}_{\mu_j,\Sigma_j}(\mathbf{O}_t)$, where $\mathcal{N}$ is a single or MVN distribution, with μ_j and Σ_j being the respective means and covariances of the different input features for state j.

In the case of the LR-based observation probabilities, as stated in Section III-G, the emission probabilities were computed using the likelihood of each state from each one-versus-all LR model. Random subsampling for each state was performed to ensure that there were a balanced number of samples in each class of the one-versus-all LR models.

In order to compare the LR-HSMM method to MVN-based observation distributions used previously by Schmidt *et al.* [23], both of these methods were tested on our dataset in order to directly compare results. Therefore, the four methods tested in this study were as follows.

1) The MVN-based method, using a single homomorphic envelope feature as described by Schmidt *et al.* [23].
2) The MVN-based method, but using the Hilbert, PSD, wavelet, and homomorphic features.

TABLE II
GROSS RESULTS OF ALGORITHMS ON TRAIN (ITALIC) AND TEST SETS (%) USING VARIOUS INPUT FEATURES, AVERAGED PER PATIENT AND ACROSS ALL 100 ITERATIONS

			Features	Viterbi Algorithm	Se	P_+	Ac	${F_1}^{S_1}$	${F_1}^{S_2}$	F_1
1	a	MVN	Homomorphic	Standard [23]	*85.09 ± 1.31*	*86.17 ± 1.27*	*77.49 ± 1.68*	*85.98 ± 1.39*	*85.34 ± 1.32*	*85.62 ± 1.29*
					85.74 ± 1.57	86.85 ± 1.53	78.35 ± 2.06	86.58 ± 1.56	86.07 ± 1.70	86.28 ± 1.55
	b			Extended	*94.29 ± 0.89*	*94.61 ± 0.86*	*90.61 ± 1.35*	*96.15 ± 0.85*	*92.73 ± 1.13*	*94.45 ± 0.88*
					94.66 ± 0.90	94.93 ± 0.88	91.15 ± 1.36	96.50 ± 0.84	93.06 ± 1.17	94.79 ± 0.89
2	a	MVN	Hilbert, PSD, Wavelet, Homomorphic	Standard	*83.08 ± 1.48*	*83.84 ± 1.45*	*74.05 ± 1.91*	*83.67 ± 1.52*	*83.32 ± 1.57*	*83.45 ± 1.47*
					83.62 ± 1.57	84.45 ± 1.53	74.82 ± 1.99	84.20 ± 1.56	83.93 ± 1.72	84.02 ± 1.55
	b			Extended	*92.74 ± 1.05*	*92.93 ± 1.04*	*87.90 ± 1.57*	*94.28 ± 1.07*	*91.37 ± 1.23*	*92.83 ± 1.04*
					93.10 ± 1.04	93.29 ± 1.03	88.46 ± 1.57	94.67 ± 1.02	91.70 ± 1.29	93.19 ± 1.04
3	a	LR	Homomorphic	Standard	*87.51 ± 1.29*	*88.72 ± 1.23*	*80.97 ± 1.68*	*87.82 ± 1.29*	*88.45 ± 1.38*	*88.10 ± 1.26*
					87.87 ± 1.34	89.10 ± 1.28	81.49 ± 1.77	88.11 ± 1.36	88.91 ± 1.43	88.47 ± 1.31
	b			Extended	*94.88 ± 0.82*	*95.28 ± 0.79*	*91.66 ± 1.25*	*96.70 ± 0.80*	*93.43 ± 1.04*	*95.08 ± 0.80*
					95.22 ± 0.88	95.60 ± 0.85	92.18 ± 1.35	97.00 ± 0.86	93.79 ± 1.12	95.40 ± 0.86
4	a	LR	Hilbert, PSD, Wavelet, Homomorphic	Standard	*87.58 ± 1.28*	*89.16 ± 1.18*	*81.22 ± 1.66*	*87.97 ± 1.32*	*88.80 ± 1.31*	*88.35 ± 1.23*
					87.84 ± 1.35	89.44 ± 1.25	81.62 ± 1.76	88.18 ± 1.37	89.13 ± 1.42	88.62 ± 1.30
	b			Extended	*95.09 ± 0.80*	*95.69 ± 0.75*	*92.11 ± 1.21*	*96.70 ± 0.84*	*94.05 ± 0.97*	*95.38 ± 0.78*
					95.34 ± 0.88	**95.92 ± 0.83**	**92.52 ± 1.33**	**96.95 ± 0.90**	**94.29 ± 1.08**	**95.63 ± 0.85**

The first line shows the previous state of the art. The last line shows the results using the proposed method (**bold**).

3) The LR-HSMM method using a single homomorphic envelope feature.
4) The LR-HSMM method using the Hilbert, PSD, wavelet, and homomorphic features.

In addition, a comparison was made between the use of the standard Viterbi algorithm and the extended Viterbi algorithm for each method listed above.

3) Model Evaluation: The four segmentation methods were evaluated on their ability to accurately locate the S_1 and S_2 sounds within the test set of recordings. The reference positions for the S_1 and S_2 sounds were the R-peaks and end-T-waves computed in Section III-B, as shown in step *vii)* in Fig. 2. An S_1 sound was labeled as correctly identified if the start of the segmented S_1 sound was found to be within 100 ms of the R-peak of the ECG. This tolerance is based on the recognized ECG R-peak detection tolerance of 150 ms [33], which, as it is approximately the length of the FHSs, was shortened to 100 ms. Similarly, an S_2 sound was labeled as correctly segmented if the center of this S_2 sound was found to be within 100 ms of the corresponding end-T-wave.

The performance of the segmentation algorithms were evaluated using the F_1 score, which is defined as

$$F_1 = \frac{2 \times P_+ \times Se}{P_+ + Se} \tag{9}$$

where Se is sensitivity (or recall) and P_+ is positive predictivity (or precision). The F_1 score was used as it gives a single harmonic mean of Se and P_+. Metrics such as accuracy (Acc $= TP/(TP + FP + FN)$) in such cases do not give an adequate representation of the results, as no true negatives are included. However, Se, P_+, and Acc are reported to allow comparison to previous works.

In order to give a robust indication of segmentation performance, the process of splitting the data into a training and test set, the wavelet feature optimization, the training of the HSMM, and the evaluation on the test set was repeated 100 times and the results averaged, as shown in step *vii)* in Fig. 2. Significance testing was then performed using a two-sided paired t-test on the 100 F_1 scores from the test datasets.

IV. RESULTS

The wavelet optimization (see Section III-H3 and step *iv)* in Fig. 2) resulted in the reverse biorthogonal 3.9 or Daubechies 10 wavelet at decomposition level 3 being selected as the optimal wavelet for each of the 100 evaluation iterations.

The gross performance results of the four algorithms under consideration on both the training and test sets, using the different features and Viterbi algorithms, are presented in Table II. This table illustrates the scores for the combined S_1 and S_2 sounds, while also presenting the $F1$ scores for each sound separately to give an indication of performance on the different sounds. These gross scores were calculated on a per patient basis, summing the total number of sounds for each patient in the train and test datasets, calculating the different metrics for each patient, then averaging over patients in each of the training and test sets. The results over the 100 iterations were then averaged. The standard deviation of the average results over the 100 evaluation iterations is also shown. The results of the current state-of-the-art algorithm [23] can be seen in the first two lines of the table (see algorithm *1a* in Table II), achieving an average F_1 score of $86.28 \pm 1.55\%$ on the test datasets. The proposed algorithm, combining the Hilbert envelope, PSD, wavelet envelope, and homomorphic envelope features along with the LR classifier and extended Viterbi algorithm, is shown in the last two lines of Table II (see algorithm *4b*), achieving an average F_1 score of $95.63 \pm 0.85\%$.

The incorporation of LR into the model alone resulted in a significant improvement of F_1 score between the current state-of-the-art algorithm (*1a*, $86.28 \pm 1.55\%$) and the LR-homomorphic algorithm with standard Viterbi algorithm (*3a*, $88.47 \pm 1.31\%$), ($p < 0.001$).

TABLE III
AVERAGE RESULTS OF ALGORITHMS $1a$ AND $4b$ FROM TABLE II WHEN AVERAGED ACROSS ALL RECORDINGS RATHER THAN PER PATIENT

		Features	Viterbi Algorithm	S_1	S_2	TP	FN	FP	$F_1(\%)$
1a	MVN	Homomorphic	Standard [23]	*6117.07*	*6012.56*	*10613.02*	*1516.61*	*1380.26*	*87.98 ± 1.06*
				5686.93	5597.44	9945.63	1338.74	1214.19	88.61 ± 1.33
4b	LR	Hilbert,PSD, Wavelet, Homomorphic	Extended (this work)	*6117.07*	*6012.56*	*11649.84*	*479.79*	*410.92*	*96.31 ± 0.58*
				5686.93	5597.44	10858.53	425.84	364.4	96.49 ± 0.64

These show the numbers and results averaged across all 100 iterations for the train (italic) and test sets. The average number of S_1 and S_2 sounds is shown, as well as the number of true positives (TP), false negatives (FN), false positives (FP) and F_1 score.

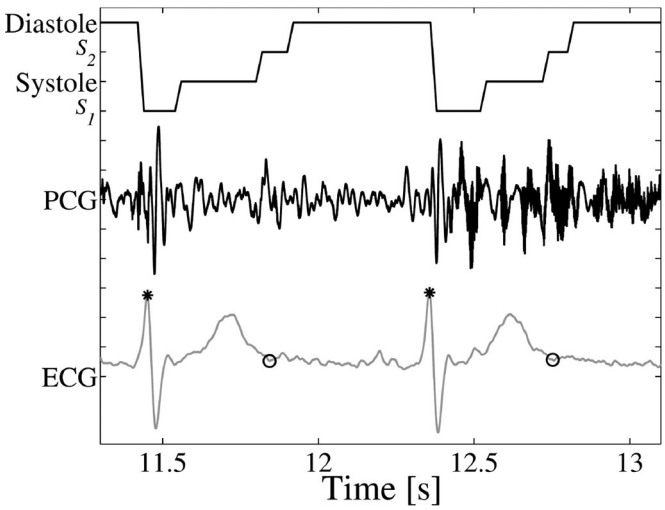

Fig. 4. Example of a segmented noisy PCG using the LR-HSMM method with a clean ECG for reference. The positions of the detected R peaks (*) and the end-T-waves (o) are demarcated in the ECG. The four LR-HSMM-derived states of the heart cycle (S_1, systole, S_2, and diastole) are shown as a solid line, which are seen to coincide with the ECG reference positions. The noise in the systolic region of the second cycle can be seen to be ignored, while the S_2 is correctly segmented.

The introduction of the Hilbert, PSD, and wavelet features, in addition to the homomorphic feature, resulted in a drop in performance when using the MVN-based HSMM (*2a*, $84.02 \pm 1.55\%$ as compared to *1a*, $86.28 \pm 1.55\%$). However, the addition of these same features resulted in a significant improvement in the performance of the LR-based HSMM ($p < 0.05$, comparing the F_1 scores of algorithms *3a* to *4a* and *3b* to *4b*).

There was a significant improvement in F_1 score when using the extended Viterbi algorithm compared to the standard Viterbi algorithm on all combinations of features and MVN or LR emission probabilities ($p < 0.001$). This resulted in at least a 6% improvement in F_1 scores across all algorithms.

The average performance results for algorithms *1a* and *4b*, averaged over all recordings in the training and test sets rather than by patient, can be seen in Table III. The average number of S_1 and S_2 sounds across the 100 iterations can be seen, as well as the performance metrics. A significant improvement on the current state-of-the-art algorithm can be seen ($88.61 \pm 1.33\%$ to $96.49 \pm 0.64\%$) ($p < 0.001$).

An illustration of the segmentation accuracy using this method on a noisy test recording is shown in Fig. 4. Note the excellent performance of the algorithm in ignoring heart sound-like noise in second systolic period, but the correct identification of the following S_2 sound which was corrupted by noise.

V. DISCUSSION

The optimal wavelets selected for all 100 evaluation iterations (see Section III-H3 and step *iv)* in Fig. 2) agree with intuition, as their center frequencies were between 65 and 85 Hz, as expected for the FHSs [41].

The results in Tables II and III, with similar results between the training and test set results, illustrate that the algorithms were not overtrained on the training data. The small values for the standard deviations in these tables indicate consistent results for the LR and MVN algorithms across the 100 evaluation iterations, with generally smaller values for the LR algorithm. This indicates that the results with the LR algorithm are more consistent than the MLR-based algorithms.

Comparing the results of the method developed by Schmidt *et al.* [23] between Tables I and II (98.8% Se and 98.6% P_+ in Table I and 85.74% Se and 86.85% P_+ in Table II) show a drop in Se and P_+ on our dataset when comparing test set results. This indicates that our data may be more noisy or heterogeneous and, therefore, more difficult to analyze. The low number of pathological cases in their dataset (8%) may also account for this.

The significant improvement in performance between the current state-of-the-art algorithm (*1a*, $86.28 \pm 1.55\%$) and the LR algorithm with the homomorphic feature and standard Viterbi algorithm (*3a*, $88.47 \pm 1.31\%$) indicates that the simple introduction of LR as opposed to the MVN-based emission probabilities significantly improves segmentation performance. This is thought to be due to the higher discrimination between states afforded by the LR model.

The introduction of the three additional features resulted in a drop in performance when using the MVN-based HSMM (comparing algorithm *1a* to *2a* and *1b* to *2b*). This was thought to be due to the MVN distribution being unable to adequately model such a higher dimensional feature space effectively. However, the introduction of these features when using the LR model resulted in a slight yet significant improvement in performance, and the best performing algorithm in this paper. This combination of features was thought to enhance the segmentation as these features contribute information about the amplitude of the

PCG, and also frequency information, in the form of the wavelet transform and PSD feature.

The largest contributor to the improved segmentation performance was the introduction of the extended Viterbi algorithm. This modification resulted in significant improvement across all algorithms in Table II (when comparing algorithm *1a* to *1b*, *2a* to *2b* and so on), where at least 6% increase in F_1 score can be seen. The weakness of the standard Viterbi algorithm was especially noted in short recordings with a short diastolic period before the onset of the first heart sound. In this case, as the standard Viterbi algorithm tried to assign a complete diastole state before the onset of S_1, the first heart sound is often missed. When using the extended Viterbi algorithm, $F_1^{S_1}$ scores were higher than $F_1^{S_2}$. This can be attributed to the patients with MVP in this dataset having systolic murmurs that often mask S_2, making it more difficult to detect.

There was an improvement in performance between the average results in Table III as compared to the gross results of Table II. The lower performance of the gross averages, which are derived by averaging per patient, indicates that specific patients (which may have fewer heart sounds as compared to others in the dataset) diminish the performance of the segmentation algorithms. This may be due to these patients having recordings contaminated by noise or having murmurs that obfuscate the positions of the FHSs.

The inclusion of explicit timing durations helps to improve the differentiation of heart sound-like noise from noisy heart sounds (see Fig. 4). Approaches which do not incorporate temporal duration and ordering information (like neural networks or other static machine learning approaches) may not perform as well in such circumstances.

VI. Conclusion

The study presented here investigated a new method for the segmentation of the S_1 and S_2 heart sounds from a single channel PCG recording with no external reference using a modified HSMM. The introduction of LR, the extended Viterbi algorithm, and additional features into the HSMM each resulted in a significant improvement in segmentation performance, the combination of which significantly improved upon the current state of the art. As demonstrated on this dataset, consisting of a large proportion of pathological recordings, this method is able to accurately segment the heart sounds in noisy real-world PCGs with murmurs and other abnormal sounds.

Acknowledgment

The authors would like to thank Dr. F. Nesta, Prof. J. Guttag, Dr. Z. Syed, and D. Curtis for the use of the auscultation data.

References

[1] A. Leatham, *Auscultation of the Heart and Phonocardiography*. London, U.K.: Churchill Livingstone, 1975.

[2] G. Douglas *et al.*, *Macleod's Clinical Examination*. New York, NY, USA: Elsevier Health Sciences, 2009.

[3] A. G. Tilkian and M. B. Conover, *Understanding Heart Sounds and Murmurs: With an Introduction to Lung Sounds*. New York, NY, USA: Elsevier Health Sciences, 2001.

[4] H. Liang *et al.*, "Heart sound segmentation algorithm based on heart sound envelogram," in *Proc. Comput. Cardiol.*, Lund, Sweden, 1997, vol. 24, pp. 105–108.

[5] H. Liang *et al.*, "A heart sound segmentation algorithm using wavelet decomposition and reconstruction," in *Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, USA, 1997, vol. 4, pp. 1630–1633.

[6] D. Kumar *et al.*, "Detection of S1 and S2 heart sounds by high frequency signatures." in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, New York, NY, USA, 2006, vol. 1, pp. 1410–1416.

[7] S. Ari *et al.*, "A robust heart sound segmentation algorithm for commonly occurring heart valve diseases." *J. Med. Eng. Technol.*, vol. 32, no. 6, pp. 456–65, Jan. 2008.

[8] J. Vepa *et al.*, "Segmentation of heart sounds using simplicity features and timing information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Las Vegas, NV, USA, 2008, pp. 469–472.

[9] C. Gupta *et al.*, "Neural network classification of homomorphic segmented heart sounds," *Appl. Soft Comput.*, vol. 7, no. 1, pp. 286–297, Jan. 2007

[10] T. Chen *et al.*, "Intelligent heartsound diagnostics on a cellphone using a hands-free kit," in *Proc. AAAI Spring Symp. Artif. Intell. Dev.*, Stanford University, Stanford, CA, USA, 2010, pp. 26–31.

[11] T. Oskiper and R. Watrous, "Detection of the first heart sound using a time-delay neural network," in *Proc. IEEE Comput. Cardiol.*, Memphis, TN, USA, 2002, pp. 537–540.

[12] A. A. Sepehri *et al.*, "A novel method for pediatric heart sound segmentation without using the ECG," *Comput. Methods Programs Biomed.*, vol. 99, no. 1, pp. 43–48, Jul. 2010.

[13] Z. Yan *et al.*, "The moment segmentation analysis of heart sound pattern," *Comput. Methods Programs Biomed.*, vol. 98, no. 2, pp. 140–50, May 2010.

[14] S. Sun *et al.*, "Automatic moment segmentation and peak detection analysis of heart sound pattern via short-time modified Hilbert transform," *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 219–230, May 2014.

[15] A. Moukadem *et al.*, "A robust heart sounds segmentation module based on S-transform," *Biomed. Signal Process. Control*, vol. 8, no. 3, pp. 273–281, May 2013.

[16] H. Tang *et al.*, "Segmentation of heart sounds based on dynamic clustering," *Biomed. Signal Process. Control*, vol. 7, no. 5, pp. 509–516, Sep. 2012.

[17] H. Naseri and M. R. Homaeinezhad, "Detection and boundary identification of phonocardiogram sounds using an expert frequency-energy based metric," *Ann. Biomed. Eng.*, vol. 41, no. 2, pp. 279–292, Feb. 2013.

[18] V. N. Varghees and K. Ramachandran, "A novel heart sound activity detection framework for automated heart sound analysis," *Biomed. Signal Process. Control*, vol. 13, pp. 174–188, Sep. 2014.

[19] C. D. Papadaniil and L. J. Hadjileontiadis, "Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1138–1152, Jul. 2014.

[20] L. Gamero and R. Watrous, "Detection of the first and second heart sound using probabilistic models," in *Proc. IEEE 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Cancun, Mexico, 2003, pp. 2877–2880.

[21] A. Ricke *et al.*, "Automatic segmentation of heart sound signals using hidden Markov models," in *Proc. Comput. Cardiol.*, Lyon, France, 2005, pp. 953–956.

[22] D. Gill *et al.*, "Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model," in *Proc. Comput. Cardiol.*, Lyon, France, 2005, pp. 957–960.

[23] S. E. Schmidt *et al.*, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, Apr. 2010.

[24] Z. Syed, "MIT automated auscultation system," Masters thesis, Dept. Electr. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2003.

[25] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, 2000.

[26] J. Behar *et al.*, "A comparison of single channel fetal ECG extraction methods," *Ann. Biomed. Eng.*, vol. 42, no. 6, pp. 1340–1353, Jun. 2014.

[27] J. Behar *et al.*, "Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data," *Physiol. Meas.*, vol. 35, no. 8, pp. 1569–1589, Aug. 2014.

[28] A. Illanes-Manriquez and Q. Zhang, "An algorithm for QRS onset and offset detection in single lead electrocardiogram records," in *Proc. IEEE 29th Annu. Int. Conf. Eng. Med. Biol. Soc.*, Lyon, France, 2007, pp. 541–544.

[29] J. P. Martínez *et al.*, "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.
[30] P. Laguna *et al.*, "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database," *Comput. Biomed. Res.*, vol. 27, pp. 45–60, 1994.
[31] Q. Zhang *et al.*, "An algorithm for robust and efficient location of T-wave ends in electrocardiograms," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12 (Pt 1), pp. 2544–2552, Dec. 2006.
[32] C. R. Vázquez-Seisdedos *et al.* (2011, Jan.). New approach for T-wave end detection on electrocardiogram: Performance in noisy conditions. *Biomed. Eng. Online* [Online]. *10(1)*, p. 77. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3201026&tool=pmcentrez&rendertype=abstract
[33] American National Standards Institute, "*Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. ANSI/AAMI Standard EC57, 2012.
[34] Q. Li *et al.*, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter," *Physiol. Meas.*, vol. 29, no. 1, pp. 15–32, Jan. 2008.
[35] J. Behar *et al.*, "ECG signal quality during arrhythmia and its application to false alarm reduction," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 6, pp. 1660–1666, Jun. 2013.
[36] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[37] N. P. Hughes, "Probabilistic models for automated ECG interval analysis," D.Phil. thesis, University of Oxford, Oxford, U.K., 2006.
[38] S.-Z. Yu, "Hidden semi-Markov models," *Artif. Intell.*, vol. 174, no. 2, pp. 215–243, Feb. 2010.
[39] F. Marzbanrad *et al.*, "Automated estimation of fetal cardiac timing events from Doppler ultrasound signal using hybrid models," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 4, pp. 1169–1177, Jul. 2014.
[40] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.
[41] P. J. Arnott *et al.*, "Spectral analysis of heart sounds: Relations between some physical characteristics and frequency spectra of first and second heart sounds in normals and hypertensives," *J. Biomed. Eng.*, vol. 6, pp. 121–128, 1984.
[42] C. Shannon, "Communication in the presence of noise," *Proc. IEEE*, vol. 86, no. 2, pp. 447–457, Feb. 1998.
[43] I. Rezek and S. Roberts, "Envelope extraction via complex homomorphic filtering," Tech. Rep. TR-98-9, Imperial College, London, U.K., 1998.
[44] S. R. Messer *et al.*, "Optimal wavelet denoising for phonocardiograms," *Microelectron. J.*, vol. 32, no. 12, pp. 931–941, Dec. 2001.
[45] D. Kumar *et al.*, "Noise detection during heart sound recording using periodicity signatures," *Physiol. Meas.*, vol. 32, no. 5, pp. 599–618, May 2011.
[46] B. Ergen *et al.*, "Time-frequency analysis of phonocardiogram signals using wavelet transform: A comparative study," *Comput. Methods Biomech. Biomed. Eng.*, vol. 15, pp. 371–381, Jan. 2011.
[47] C. Ahlstrom *et al.*, "A method for accurate localization of the first heart sound and possible applications," *Physiol. Meas.*, vol. 29, no. 3, pp. 417–428, Mar. 2008.
[48] H. Uguz *et al.*, "A biomedical system based on hidden Markov model for diagnosis of the heart valve diseases," *Pattern Recog. Lett.*, vol. 28, no. 4, pp. 395–404, Mar. 2007.
[49] S. Choi, "Detection of valvular heart disorders using wavelet packet decomposition and support vector machine," *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1679–1687, Nov. 2008.

Authors' photographs and biographies not available at the time of publication.