A MULTISCALE MEAN SHIFT LOCALIZATION APPROACH FOR ROBUST EXTRACTION OF HEART SOUNDS IN RESPIRATORY SIGNALS

Feng Jin[1] *and Farook Sattar*[2]

[1]Dept. of Electrical & Computer Engineering, Ryerson University, Toronto, M5B 2K3, Canada.
[2]Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

ABSTRACT

This paper addresses the problem of heart sound (HS) extraction in different types of single-channel respiratory sound (RS) signals by proposing a multiscale mean shift localization approach. First, the incoming respiratory signal (RS) are identified into linear/nonlinear portions by using third-order cumulant. Second, the identified linear and nonlinear portions are processed separately to tackle the large variations in the signal characteristics of adventitious sounds. The time-varying mean-shifts of the weighted log likelihood ratios of wavelet features are then calculated to capture the signal dynamics of various noisy RS signals. The proposed approach provides promising results giving an overall false localization rate as low as $(1.8 \pm 1.8)\%$ for normal lung sound (LS) and $(0.1 \pm 1.7)\%$ for adventitious sound signals. Therefore, the presented approach successfully attempts to solve the key clinical challenges faced by the existing localization methods in terms of respiratory ailments.

***Index Terms*—** Multiscale Mean Shift Localization, Heart Sound (HS), Respiratory Sound (RS), Adventitious Sound, Multiscale Decomposition.

1. INTRODUCTION

The localization of HS in normal RS signal is usually achieved by thresholding the extracted feature sequence based on some selected properties of RS and HS signals. Adaptive thresholding has been the most practical approach as it does not require any *a priori* information of the signal. Then certainly the accuracy of HS localization depends on both thresholds as well as effectiveness of the extracted features by discriminating HS from RS. In the literature, the thresholds have been calculated either globally or segment-wise based on extracted features such as average power [1][2] [3], lower order statistics of the original signal [4], lower order statistics of feature domain values [5]. On the other hand, the differentiating features of HS and RS also include the time-frequency domain power differences [1], signal singularities [4], signal distribution [5], physiological dynamics [6] and signal complexity [7]. The HS localization methods based on the above mentioned discriminating features have been summarized in [8] and some comparison results are listed in [9]. Recently, another method is proposed for localizing HS in respiratory signals using singular spectrum analysis (SSA) based on an effective time series analysis technique [10].

All of the above mentioned techniques have been developed for RS signals captured over the chest (i.e. lung sound (LS)). However, besides LS, tracheal breath sound (TBS) being a RS signal detected over the extra thoracic portion of trachea, can be more appropriate for the analysis of pathological sounds originating in larynx or trachea (such as stridor). Therefore, real TBS recordings are also considered in this paper to evaluate the performance of our HS localization framework. In fact, TBS is characterized by its broader noise spectrum containing higher frequencies compared to the LS [11] signal. Because of that the spectra of HS signals captured over the suprasternal notch are wide compared to the spectra of HS signals recorded over the chest. It can be noted that since our focus here is HS localization for different types of continuous adventitious sounds rather than normal breath sound, the large variations to the types of input signals as well as their amplitudes certainly downgrade the performance of the existing methods based on power, dynamics, complexity and distribution of the signal.

Hence, the aim of this paper is to develop a HS localization scheme to be effective for different types of RS signals. The proposed localization approach is based on multiscale mean-shift of the weighted log likelihood ratio of the wavelet features. The performance of the method is evaluated in terms of false detection rate and boundary localization accuracy for real RS signals. Moreover, the proposed approach is compared to a recent Shannon entropy based method in [5] being chosen for comparison due to its reliable performance under different amplitude ratios of HS and LS. The originality of the proposed method compared to the existing methods for HS localization lies in: (1) multiscale mean-shift approach for log likelihood ratios of the wavelet features; (2) two-stage approach consists of identification of linear/nonlinear portions of input RS signal followed by HS localization based on time-varying mean-shift which seems to be robust to various types of RS signals.

2. METHODOLOGY

2.1. Signal Model

In particular, RS as heard over the large airways is primarily related to the vibrations of the upper airway walls and turbulent airflow, while HS occurs mainly due to the valvular activity of the heart. Both hypothetical sources of HS and RS can be approximately considered as point sources [12] which are assumed to be mutually uncorrelated. However, the assumption is no longer valid at the recording position since both RS and HS share some common transmission path before recorded at the suprasternal notch.

The noisy input signal $\mathbf{y}$ is considered here to be a continuous RS signal of different types contaminated by discontinuous HS signals. Then a corresponding noisy segment corrupted by the jth HS signal is denoted by $\mathbf{y}_j$, which can be expressed as

$$\mathbf{y}_j = \mathbf{h}(\mathbf{s}_j + \mathbf{v}_j), \quad 1 \leq j \leq J \tag{1}$$

where $\mathbf{s}_j$ and $\mathbf{v}_j$ are $(N \times 1)$ vectors denoting underlying RS component of different types and HS component respectively. The total number of occurrence of HS signal is represented by J. The $\mathbf{v}_j$ or the jth occurred HS signal, which has a transient waveform, is then superimposed onto the uncorrelated RS signal $\mathbf{s}_j$, whereas $\mathbf{h}(\cdot)$ represents the effect of transmission path through trachea and skin. In this paper, we assume additive signal model of RS and HS, however do not consider the possibilities of multiplicative model, or even more complex interaction for our initial work presented in the following.

2.2. Proposed HS Localization Approach

2.2.1. Identification of Nonlinear Portions Based on Third-order Cumulants

In this section, third-order cumulant is used in order to identify the nonlinear portions of the input signal as derived in our previous work, see [13] for more details. The nonlinear portions of the input signal are extracted partitioning the rest being linear portions so that

$$\mathbf{y} = \| \sum_{g=1\ldots\mathbf{G}} \left\{ \mathbf{y}_g^L, \mathbf{y}_g^{NL} \right\} \tag{2}$$

where $\mathbf{y}_g^L$ and $\mathbf{y}_g^{NL}$ refer to the gth linear and nonlinear portions of $\mathbf{y}$, with $\mathbf{G} = \{G_1, G_2\}$ referring the total number of identified linear and nonlinear portions. Incidentally, if the input noisy RS signal has $\max(\rho_3) \leq 10^{-8}$, then $G = 0$ considering that the RS signal only has linear portions. The identified linear/nonlinear portions are then concatenated alternatively as defined by the symbol $\|\sum$ (e.g. $\mathbf{y} = \underset{g=1,...,\mathbf{G}}{\|\sum}\{\mathbf{y}_g^L, \mathbf{y}_g^{NL}\} = [\mathbf{y}_1^L; \mathbf{y}_1^{NL}; \cdots; \mathbf{y}_g^L; \mathbf{y}_g^{NL}; \cdots \mathbf{y}_G^{NL}; \mathbf{y}_G^L]$ if the signal starts and ends with linear portions). Fig. 1(a-b)(left panel) show illustrative plots of the third-order cumulant and the corresponding identified linear/nonlinear portions of a real RS signal. The HS locations $\mathbf{A}_g^L$ and $\mathbf{A}_g^{NL}$ obtained in the following section 2.2.2 will be concatenated in a similar way as shown in (2).

The choice of window length to estimate the cumulant in [13] is crucial in our proposed approach. It should be selected properly in order to able to track the rapid variations while ignoring relative slow variations. Here, window length of $2B + 1$ is selected to be much less than minimum duration of HS signal as $1/20$ of the minimum duration of HS signal (which is 20 ms [17]), for which we have the acceptable range $10 \leq B \leq 21$ for $Fs = 44.1$ kHz, while $B \leq 10$ produces some distortion due to overestimation. It is worthy to mention that the respiratory system can be considered as an acoustic system where the acoustic properties can be calculated using electro-acoustic analogy [14]. And the linear/nonlinear components of respiratory system [15] basically originate the linear/nonlinear portions of HS signals.

2.2.2. Multiscale Decomposition

The multiscale analysis of the proposed algorithm is done by time-scale decomposition of an input signals by wavelet transform. This affine time-scale transformation of the identified $\mathbf{y}_g^L$ or $\mathbf{y}_g^{NL}$ then take the following matrix form:

$$\mathbf{y}_n = \mathbf{W}_n^H \mathbf{y}_p \tag{3}$$

where H is conjugate transpose. $\mathbf{y}_n$ is with size $(M \times 1)$, $\mathbf{y}_p$ refers to $\mathbf{y}_g^L$ or $\mathbf{y}_g^{NL}$ with size $(N_p \times 1)$, and $\mathbf{W}_n$ is a $(N_p \times M)$ matrix given by

$$\mathbf{y}_n = [y(n,0) \;\; y(n,1) \cdots y(n, M-1)]^T$$
$$\mathbf{y}_p = [y(n) \;\; y(n-1) \cdots y(n-N_p+1)]^T$$

$$\mathbf{W}_n = \begin{bmatrix} \psi(n,0) & \cdots & \psi(n, M-1) \\ \psi(n-1,0) & \cdots & \psi(n-1, M-1) \\ \vdots & \cdots & \vdots \\ \psi(n-N_p+1, 0) & \cdots & \psi(n-N_p+1, M-1) \end{bmatrix}$$

with $(\cdot)^T$ denoting the transposition and $y(n,m)$ representing the wavelet transform coefficient of $\mathbf{y}_p$ at the mth scale. Among different choices in the wavelet family, Morlet wavelet is employed here because of its ability to provide different window lengths for signals composed of different frequencies/scales. The Morlet wavelets form non-orthogonal basis functions with considerable spectral overlap among them and can be expressed by wavelet functions $\{\psi(n,m) = \pi^{-1/4} e^{j2\pi f_o n/m} e^{-n^2/2m^2}\}$, with $n = 0, \cdots, N_p - 1$ being the time indices and $m = 0, \cdots, M-1$ being the scale indices. N_p refers to the total number of samples of the noisy signal $\mathbf{y}_g^L$ or $\mathbf{y}_g^{NL}$. In compare to discrete wavelet transform, the continuous wavelet transform being shift invariant, is able to provide high HS localization accuracy.

2.2.3. Proposed Multiscale Mean Shift Localization

In the following, we propose a multiscale mean shift localization for extraction of heart sounds in various respiratory signals. The localization of HS signal is initiated by proposing time-varying mean-shift that is proportional to the sum of the log likelihood ratios of the multiscale features for the two hypothesis, H_0 and H_1.

$$H_0 : \mathbf{y}_n = \mathbf{s}_n; \quad H_1 : \mathbf{y}_n = \mathbf{s}_n + \mathbf{v}_n \tag{4}$$

where $\mathbf{s}_n$, $\mathbf{v}_n$ and $\mathbf{y}_n$ are the wavelet transform coefficients of $s(n)$, $v(n)$ and $y_p(n)$, respectively. At a given time instance n, we have to decide if $\mathbf{y}_n$ contains the HS component $\mathbf{v}_n$, or it contains only the RS component $\mathbf{s}_n$.

We employ the maximum likelihood approach, which provides Generalized Log Likelihood Ratio (GLR) for the noisy input sequence $\mathbf{y}_n$ and the underlying RS component $\mathbf{s}_n$ with unknown covariances given by [18]

$$L(\mathbf{y}_n) = ln\ p(\mathbf{y}_n, \hat{\mathbf{R}}_{yy}; H_1) - ln\ p(\mathbf{y}_n, \hat{\mathbf{R}}_{ss}; H_0) \tag{5}$$

where $L(\cdot)$ is the log likelihood ratio with $\hat{\mathbf{R}}_{yy}$, $\hat{\mathbf{R}}_{ss}$ as the maximum likelihood estimates of the covariance under hypothesis H_1 and H_0, respectively. $\mathbf{s}_n$ and $\mathbf{v}_n$ are uncorrelated for the hypothesis in (4).

For this, the covariance matrix of $\mathbf{y}_n$, $\mathbf{R}_{yy}$, under hypothesis H_1 is determined by

$$\mathbf{R}_{yy} = E[\mathbf{y}_n \mathbf{y}_n^H] \tag{6}$$

with $E[\cdot]$ being the expectation operator. Using SVD of the covariance matrix $\mathbf{R}_{yy}$, we get $\mathbf{R}_{yy} = \mathbf{U\Lambda U}^T$, $\mathbf{\Lambda} = diag[\sigma_1^2 \cdots \sigma_I^2]$ where $\mathbf{U}$ is the eigenvector matrix of $\mathbf{R}_{yy}$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_I$ are the corresponding singular values with I denotes the total number of singular values.

Then the log-conditional probability density function of $\mathbf{y}_n$ under H_1 is

$$lnp(\mathbf{y}_n; H_1) = -\frac{I}{2} ln 2\pi - \frac{1}{2} \sum_{i=1}^{I} \left(ln(\sigma_i^2) + \frac{\mathbf{y}_n^H U_i U_i^H \mathbf{y}_n}{\sigma_i^2} \right). \tag{7}$$

The log-conditional probability density function of $\mathbf{y}_n$ under the hypothesis H_0 becomes

$$lnp(\mathbf{y}_n; H_0) = -\frac{I}{2} ln 2\pi - \frac{I}{2} ln \sigma_s^2 - \frac{\mathbf{y}_n^H \mathbf{y}_n}{\sigma_s^2} \tag{8}$$

Using (7) and (8), we get

$$L(\mathbf{y}_n) = -\frac{1}{2}\sum_{i=1}^{I} ln\sigma_i^2 + \frac{I}{2} ln\sigma_s^2 + \frac{1}{2\sigma_s^2}\sum_{i=1}^{I}\left(\frac{\sigma_i^2-\sigma_s^2}{\sigma_i^2}\right)|U_i\mathbf{y}_n|^2. \tag{9}$$

In order to obtain discriminant features, feature selection is considered here by using linear decomposition based on SVD and selecting the frequency scales. Since the covariance matrix can be approximated as $\hat{\mathbf{R}}_{yy} = \hat{\mathbf{R}}_{ss} + \hat{\mathbf{R}}_{vv}$ and $\hat{\mathbf{R}}_{yy} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^H; \hat{\mathbf{\Lambda}} = diag[\hat{\sigma}_1^2, \hat{\sigma}_2^2, \cdots, \hat{\sigma}_I^2]$, the SVD of $\hat{\mathbf{R}}_{ss}$ and $\hat{\mathbf{R}}_{vv}$ can therefore be expressed in terms of the singular matrix $\hat{\mathbf{U}}$ of $\hat{\mathbf{R}}_{yy}$ as

$$\begin{array}{ll} \hat{\mathbf{R}}_{vv} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}_{vv}\hat{\mathbf{U}}^H & \hat{\mathbf{\Lambda}}_{vv} = diag[\hat{\sigma}_{v1}^2, \hat{\sigma}_{v2}^2, \cdots, \hat{\sigma}_{vI}^2] \\ \hat{\mathbf{R}}_{ss} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}_s\hat{\mathbf{U}}^H & \hat{\mathbf{\Lambda}}_s = \hat{\sigma}_s^2.I \\ \hat{\mathbf{\Lambda}} = \hat{\mathbf{\Lambda}}_v + \hat{\mathbf{\Lambda}}_s & \hat{\sigma}_i^2 = \hat{\sigma}_{vi}^2 + \hat{\sigma}_s^2. \end{array} \tag{10}$$

By separating the dominant singular values from the rest as

$$\hat{\mathbf{\Lambda}} = diag[\hat{\sigma}_1^2, \hat{\sigma}_2^2, \cdots, \hat{\sigma}_r^2, \hat{\sigma}_{r+1}^2, \cdots, \hat{\sigma}_I^2] \tag{11}$$

with $\hat{\sigma}_{\{i|i=1,2,\cdots,r\}}$ being the dominant or principal singular values, the singular matrix $\hat{\mathbf{U}}$ can be partitioned into

$$\hat{\mathbf{U}} = [\hat{\mathbf{U}}^v \quad \hat{\mathbf{U}}^s], \quad \hat{\mathbf{U}}^v = [\hat{\mathbf{U}}_1 \cdots \hat{\mathbf{U}}_r], \quad \hat{\mathbf{U}}^s = [\hat{\mathbf{U}}_{r+1} \cdots \hat{\mathbf{U}}_I].$$

The dominant singular values $\hat{\sigma}_{\{i|i=1,2,\cdots,r\}}$ and the corresponding dominant singular matrix $\hat{\mathbf{U}}^v$ characterize the signal plus noise components if $\hat{\sigma}_{\{i|i=1,2,\cdots,r\}} >> \hat{\sigma}_s$. The columns of the $\hat{\mathbf{U}}^v$ matrix therefore spans the signal plus noise space, while the columns of the $\hat{\mathbf{U}}^s$ spans its orthogonal complement, namely the noise space [16]. The singular matrix $\hat{\mathbf{U}}$ under the hypothesis H_1 can be then substituted by its estimate $\hat{\mathbf{U}}^v$. At the same time, the singular values $\{\sigma_i\}$ and $\hat{\sigma}_s$ can be substituted by their estimates $\hat{\sigma}_{\{i|i=1,2,\cdots,r\}}$ and $\hat{\sigma}_{\{i|i=r+1,\cdots,I\}}$ respectively. Thus, $\hat{\mathbf{U}} \approx \hat{\mathbf{U}}^v$, $\hat{\sigma}_i \approx \hat{\sigma}_{\{i|i=1,2,\cdots,r\}}$, and $\hat{\sigma}_s \approx \frac{1}{I-r}\sum_{i=r+1}^{I}\hat{\sigma}_i$.

Our localization approach is then originated from the idea of a time-varying mean-shift $d(n)$ based on the weighted log likelihood ratios of the selected wavelet features defined as

$$d(n) = \left(\sum_{n-N_d-1}^{n}\frac{\rho_n L(\mathbf{y}_n)}{\sum_{n-N_d-1}^{n}\rho_n}\right) - L(\mathbf{y}_n) \tag{12}$$

$L(\mathbf{y}_n)$ is calculated using (5), where the log likelihood of HS, $ln\ p(\mathbf{y}_n, \hat{\mathbf{R}}_{yy}; H_1)$, is calculated using (6) with coefficients $\mathbf{y}_n$ from scale 1 to 2. Similarly, the log likelihood of the underlying RS component, $ln\ p(\mathbf{y}_n, \hat{\mathbf{R}}_{ss}; H_0)$, is found by using $\mathbf{y}_n$ from scales 3 to 24. Since scale 2 covers the whole frequency range of HS component, the remaining scales can be considered for true RS. Furthermore, the weighting function ρ_n is computed as the geometric average of the wavelet coefficients along the scales to give more importance to high-energy points and less to low-energy points.

The choice of N_d is based on the minimum duration of HS signal so that the window length is more than the length of the HS signal covering the target object. Since the input signal is divided into 24 subbands(as it consists of 8 uniform bands followed by decomposing each band into 3 octave bands), the minimum duration of HS signal is thereby scaled from 20 ms to 5 ms (since sampling rate F_s is reduced by 4 due to the generation of 3 octave bands). $N_d \leq 200$ is therefore found to be an acceptable approximate range with respect to minimum duration of HS signal and number of octave bands for F_s=44.1 kHz calculated as $N_d = (20/1000) * (F_s/2^{(3-1)})$, from which we approximately set N_d =200.

Lastly a set of samples corresponding to large $d(n)$ is obtained to extract the HS signals based on the following selected thresholds.

Threshold Selection

Since $d(n)$ captures the signal dynamics corresponding to TBS signal, a threshold γ_e

$$\gamma_e = \frac{1}{N_p}\sum_{n=1}^{N_p} d(n) \tag{13}$$

with N_p being the number of samples for the noisy signal $\mathbf{y}_p$, is applied to $d(n)$ to get rid of noise effect and extract HS signals. Finally, in order to improve the false and correct localization rates, temporal thresholding is applied based on the following rules:

- Located segments smaller than 20 ms are discarded based on the minimum duration of HS signals. [17].
- Located segments within 50 ms are merged based on the maximum interval of normal HS signals [17].

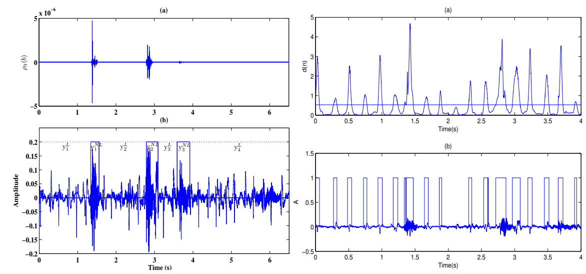

Fig. 1. (a) Left: Third-order cumulant of a real RS signal; (b) Original waveform of the RS signal with its respective linear (dashed line)/nonlinear (solid line) portions being labeled manually; Right: (a) Time-varying mean-shift $\mathbf{d}$ for the noisy signal $\mathbf{y}$ displayed in (c); (b) Noisy RS signal together with HS locations $\mathbf{A}$.

3. RESULTS AND DISCUSSION

3.1. Data

The synthesized data are generated by superimposing true HS recording onto various types of standard RS signal of same length as provided by [19] and [20]. Five different types of RS signals including normal LS for adults and infants, expiratory mild wheeze, inspiratory stridor and expiratory moderate wheeze, monophonic wheeze, and polyphonic wheeze, each of 10 seconds duration, have been used for our simulation. The true HS recordings (which are phonocardiogram (PCG) signals) together with synchronized electrocardiogram (ECG) signals for healthy subjects are acquired from the Department of Cardiology, Lund University Hospital, Sweden, with permission.

3.2. Performance Index

The performance of the proposed HS localization approach is measured as a percentage of “true” S1 and S2 activities that have been “accurately” located. A fundamental activity S1 has been counted as accurately located with estimation error $\epsilon = 0\%$ if the labeled S1 region of the synthesized signal coincides with the peak n the QRS complex in its synchronously recorded ECG signal. On the other hand, correct S2 locations with estimation error $\epsilon = 0\%$ are decided by the experienced doctors through listening and labeling of true HS signals. Any misalignment in the simulated signal increases the estimation error ϵ as defined by

$$\begin{cases} \epsilon_j &= \frac{1}{2}\left\{ \left| \frac{\widehat{P}_{Sj} - P_{Sj}}{D_j} \right| + \left| \frac{\widehat{P}_{Ej} - P_{Ej}}{D_j} \right| \right\} \\ \mu &= \sum_{j=1}^{J} \epsilon_j \\ \epsilon &= \mu \pm \sqrt{\sum_{j=1}^{J} (\epsilon_j - \mu)^2} \end{cases} \quad (14)$$

ϵ_j is the percentage error of the jth located HS segments, with $\widehat{P}_{Sj}$ and $\widehat{P}_{Ej}$ being the estimated starting and end positions of the jth HS segment obtained from the HS location sequence $\mathbf{A}$. P_{Sj} and P_{Ej} are the benchmark start and end positions of the jth HS segment with D_j being the duration of the jth HS segment obtained based on the doctors' decision. J represents the total number of HS segments.

3.3. Results

Fig. 2 illustrates the HS localization results for a synthetic noisy mild wheeze signal for the proposed method and the entropy based method in [5]. Synchronized ECG signal is also displayed in Fig. 2(c) for reference. The performance comparison of these two methods for different types of HS contaminated RS signals is summarized in Table 1. Mean and standard deviation of the estimation error between the actual HS locations (based on synchronized ECG signals and doctors' decision) and the estimated HS locations for the synthesized noisy RS signals are calculated. For performance evaluation, the ϵ is calculated for each subject using (14) followed by averaging over the subjects. The proposed approach gives an overall

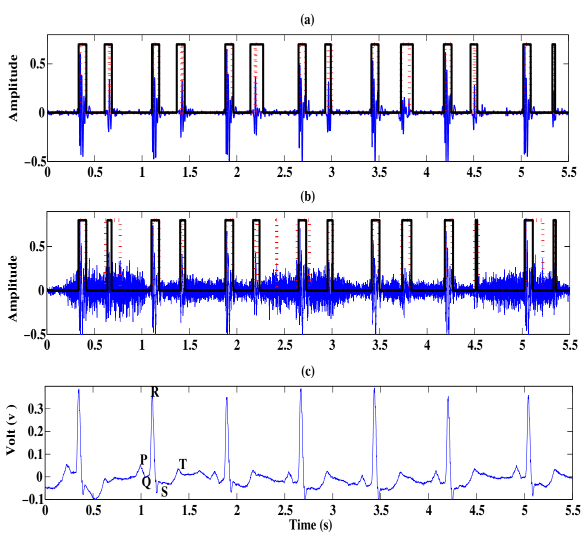

Fig. 2. (a) HS localization results for the true HS signal using the proposed method (solid line) and the entropy based method in [5] (dotted line); (b) HS localization results for the synthesized noisy mild wheeze signal using the proposed method (solid line) and the entropy based method (dotted line); (c) The synchronized ECG signal.

estimation error as low as $(0.1 \pm 2.5)\%$ for the true HS signals as listed in Table 1. Furthermore, due to the different signal characteristics, the performance of the presented method is significantly better for wheeze and stridor than that of normal LS signal. According to the definition in [11], both stridor and wheeze are categorized as continuous adventitious sound (CAS) which are characterized by their periodic waveforms and dominant spectral components over 100 Hz, while normal LS is characterized by broad spectral noise. It seems that the temporal variations are much less for CAS signals than normal LS signals. This makes the difference in $d(n)$ values (see Eq. (12)) between CAS signals with and without HS signals to be larger and it results in more accurate HS localization for the slow-varying adventitious signals.

Except for moderate wheeze, it happens for all other types of signals that the distributions and the standard deviations between the RS signals with and without HS are quite similar which downgrades the performance of the entropy based method in [5]. Therefore the proposed approach outperforms the entropy based method in all signal conditions except the moderate wheeze (see Table 1).

Table 1. Comparison of the localization accuracy for different HS localization methods in terms of the estimation error ϵ (%)

Type of Signal	**The proposed method**	**The method in [5]**
True HS	0.1 ± 2.50	2.1 ± 1.60
Normal Adult LS	1.8 ± 1.80	2.0 ± 0.87
Normal Infant LS	1.9 ± 5.43	2.3 ± 2.46
Stridor with Mild Wheeze	0.1 ± 1.70	1.7 ± 4.68
Stridor with Moderate Wheeze	1.7 ± 1.05	0.8 ± 1.05
Polyphonic Wheeze	1.2 ± 2.06	2.6 ± 7.62
Monophonic Wheeze	0.6 ± 1.08	0.55 ± 1.21

4. CONCLUSION

This paper proposes a new approach for HS localization in different types of single-channel RS signals based on an idea of time-varying mean-shift. A multiscale mean shift function using the weighted log likelihood ratios of selected wavelet features is proposed here capturing the temporal dynamics of input RS signals. The performance evaluated for synthesized real LS data and TBS data to validate the proposed method to different types of RS signals shows promising results. Since none of the existing HS localization methods works well for different types of RS signals, the proposed approach is a successful first attempt to open the exploration into a greater field. The approach described in this paper is, at the best of our knowledge, the first approach based on the idea of mean shift localization for extracting HS in RS signals. However, the challenges remain for our future work to localize abnormal HS (e.g. murmurs) in the presence of discontinuous adventitious sounds (such as crackles) due to its different dynamic nature.

Acknowledgment
The authors gratefully acknowledge the contribution of National University Hospital(NUH), Singapore, especially Dr. D.Y.T. Goh and Dr. Irene Melinda Louis, for their support in data collection and identification.

5. REFERENCES

[1] M. T. Pourazad, Z. Moussavi, and G. Thomas, “Heart sound cancellation from lung sound recordings using time-frequency filtering”, *Med. Biol. Eng. Comput.*, vol. 44, no. 3, 2006, pp. 216–225.

[2] M. Kompis and E. Russi, “Adaptive heart-noise reduction of lung sounds recorded by a single microphone”, *Proc. 14th IEEE Eng. Med. Biol. Soc. (EMBS)*, vol. 2, 1992, pp. 691–692.

[3] L. J. Hadjileontiadis and S. M. Panas, “Adaptive reduction of heart sounds from lung sounds using fourth-order statistics”, *IEEE Trans. Biomed. Eng.*, vol. 44, no. 7, 1997, pp. 642–648.

[4] D. Flores-Tapia, Z. Moussavi, and G. Thomas, “Heart sound cancellation based on multiscale products and linear prediction”, *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, 2007, pp. 234–243.

[5] A. Yadollahi and Z. Moussavi, “A Robust method for heart sounds localization using lung sounds entropy”, *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, 2006, pp. 497–502.

[6] C. Ahlstrom, O. Liljefeldt, P. Hult, and P. Ask, “Heart sound cancellation from lung sound recordings using recurrence time statistics and nonlinear prediction”, *IEEE Signal Process. Letters*, vol . 12, no. 12, 2005, pp. 812–815.

[7] J. Gnitecki and Z. Moussavi, “Variance fractal dimension trajectory as a tool for heart sound localization in lung sounds recordings”, *Proc. IEEE Eng. Med. Biol. Soc. (EMBS)*, vol. 3, no. 3, 2003, pp. 2420–2423.

[8] J. Gnitecki and Z. Moussavi, “Separating heart sounds from lung sounds”, *IEEE Eng. Med. Bio. Mag.*, vol. 26, no. 1, 2007, pp. 20–29.

[9] Z. Moussavi, “Fundamentals of Respiratory Sounds and Analysis”, *Synthesis Lectures on Biomedical Engineering*, Ed. J. D. Enderle, Morgan & Claypool Publishers, 2007.

[10] F. Ghaderi, HR. Mohseni, and S. Sanei,“Localizing heart sounds in respiratory signals using singular spectrum analysis”, *IEEE Trans Biomed Eng.*, 58(12), 2011, pp. 3360–3367.

[11] A. R. A Sovijärvi, L. P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, and J. E. Earis, “Characteristics of breath sounds and adventitious respiratory sounds”, *Eur Respir Rev*, vol. 10, no. 77, 2000, pp. 591–596.

[12] M. Kompis, H. Pasterkamp, Y. Motai, and G. R. Wodicka, “Spatial representation of thoracic sounds”, *Proc. IEEE Eng. Med. Biol. Soc. (EMBS)*, vol. 20, no. 3, 1998, pp. 1661–1664.

[13] F. Jin and F. Sattar, “Tracking and time-frequency analysis on nonlinearity of tracheal sounds”, *Medical and Biological Engineering and Computing*, Vol. 47, No. 4, 2009, pp. 457–461.

[14] M. Razanek and K. Roubik, “Mathematical model of the respiratory system– Comparison of the total lung impedance in the adult and neonatal lung”, *Int. Journal of Biological and Life Sciences*, 2(4), 2006, pp. 250–253.

[15] E. Saatci and A. Akan, “Inverse modeling of respiratory system during noninvasive ventilation by maximum likelihood estimation”, *Eurasip Journal on Advances in Signal Processing*, vol. 2010, 2010.

[16] A. -J. Van Der Veen, E. F. Deprettere, and A. L. Swindlehurst, “Sub-space based signal analysis using singular value decomposition”, *Proc. IEEE*, vol. 81, no. 9, 1993, pp. 1277–1307.

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

[18] L. L. Scharf, “Statistical Signal Processing: Detection, Estimation, and Time Series Analysis”, Addison-Wesley, 1991.

[19] S. Lehrer, “Understanding Lung Sounds”, Philadelphia, Audio CD, 2002.

[20] R. L. Wilkins, J. E. Hodgkin, and B. Lopez, “Fundamentals of Lung and Heart Sounds”, Mosby, Audio CD, 2004.