LOG-FREQUENCY SPECTROGRAM FOR RESPIRATORY SOUND MONITORING

Feng Jin[1]*, Farook Sattar*[2]*, and Sridhar Krishnan*[1]

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

ABSTRACT

Computerized patient monitoring provides valuable information on clinical disorders in medical practice, and it triggers the need to simplify the extent of resources required to describe large set of complex biomedical signals. In this paper, we present a new signal quantification method based on block-wise similarity measurement between the neighboring regions in the optimized log-frequency spectrogram of audio signals. Low dimensional cepstral feature set for signal quantification is then formed from the reconstructed similarity matrix using 2D principal component analysis. The effectiveness of the method is verified with real respiratory sound (RS) signals for the purpose of abnormal RS detection towards RS monitoring. Unlike conventional pathological RS detection methods which extract features from well-segmented inspiratory/expiratory phase segments, the proposed scheme is able to perform fast detection of various types of abnormality for unsegmented signals.

Index Terms— Log-Frequency Spectrogram, Mutual Information, 2D Principal Component Analysis, Respiratory Sound Monitoring, Signal Quantification

1. INTRODUCTION

Respiratory sound (RS) signals carry significant information about the underlying functioning of pulmonary system by the presence of adventitious sounds. As proposed by American Thoracic Society in 1980 [1], a general classification scheme of RS has been applied to divide RS into normal and abnormal sounds, with the latter being further divided as continuous adventitious sounds (CAS) and discontinuous adventitious sounds (DAS). According to [2], normal RS were characterized by broadband noise. On the other hand, CASs were quasi-stationary signals longer than 100 ms with various fundamental frequencies and DASs were transient signals shorter than 20ms with a wide frequency band $[50, 2000]$ Hz.

In literature, RS signals have been analyzed in joint time-frequency (TF) domain to accommodate their inherent non-stationarity at a moderately low computational complexity. Many studies have addressed the problem of pathological RS detection from a spectral-temporal analysis stage by quantifying the respective TF structures related to various types of RS based on the definition discussed above. The success of signal quantification based on these features for abnormality detection relies on the accurate respiratory phase segmentation[3], the adoption of extra decision rules based on signal characteristics [4], as well as the resolution of the TF representation [5]. It is time consuming to apply segment-based method for long-term patient monitoring and the adoption of decision rules based on various RS signal characteristics reduce the robustness of the method. On the other hand, only a few scientific works[6] have focused on unsegmented signal analysis. However, quasi-stationarity of CAS signals together with a minimum requirement of signal length were adopted that naturally limiting the detection of DAS (being transient in nature) and the possibility of online abnormal RS detection towards RS monitoring.

In this paper, we propose a reliable signal quantification method based on the construction of an optimized log-frequency (Logf) spectrogram, and demonstrate its application on real RS recordings. By mapping the short-time Fourier transform (STFT) of the signal into Logf scale using a weighting function, without identifying the characteristic spectral-temporal structure of various sounds, the discriminative features are solely selected based on capturing the non-specific signal variability at each spectral-temporal instance by treating the TFR of the signal as a image. Similarities between each pair of adjacent subregions in the TFR are computed to quantify the respective TF structure of each signal, followed by 2DPCA reconstruction for feature selection. A new signal quantification that supports immediate binary clustering is lastly obtained from the reconstructed image plane and thus enables automatic abnormal RS detection from unsegmented RS signals.

2. METHODOLOGY

In this section, a signal quantification approach is proposed to output a discriminative compact feature set for audio signals annotation with the strategy being summarized in Fig. 1.

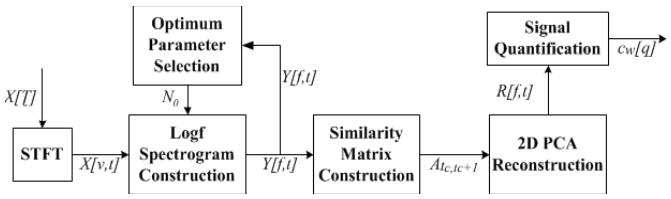

Fig. 1. Block diagram of the proposed signal quantification method.

2.1. Log-frequency Spectrogram

The log-frequency (Logf) spectrogram is generated as a mapping applied to a STFT representation. Each bin in the Logf spectrogram is formed as a weighting of corresponding frequency bins from the original spectrogram. For a Logf axis the calculation can be expressed in matrix notation as $Y{=}HX$, where Y is the Logf spectrogram, X is the original STFT with magnitude $|X[\nu,t]|$ (with ν indexing frequency and t indexing time). H is the weighting matrix, being a Gaussian function here, that gives the weight of STFT bin $|X[\nu,.]|$ contributing to Logf bin $Y[f,.]$.

$$H[f,\nu] = exp\left\{-\frac{1}{2B^2}\left(log_2\frac{\nu}{\nu_{min}} - \frac{f}{N_o}\right)^2\right\} \qquad (1)$$

where B defines the bandwidth as the frequency difference (in octaves) at which the bin has fallen to $exp(-1/2)$ of its peak gain. ν_{min} is the frequency of the lowest bin ($f{=}0$), N_0 is the number of bins per octave in the log-frequency axis. Weighting matrix mapping the energy in FFT bins to Logf bins is a set of Gaussian profiles depending on the difference in frequencies, scaled by the bandwidth of that bin. Here we have used a time window of 1024 samples, or almost 23 ms at $F_s{=}44.1$ kHz to achieve semitone resolution at low frequencies. The Log axis uses $\nu_{min}{=}200$ Hz to suppress heart sound interference while maintaining most of the signal energy [2].

The optimum parameter selection approach, which refers to optimum choice of N_o is based on the consideration that at optimal N_o, the response of the H function will produce sharp contour features with highest average intensity, while for other N_o values the average intensity will be smaller. In order to measure the contour intensity, we construct the Hessian matrix, Γ at point (f,t) defined as

$$\Gamma = \begin{bmatrix} L_{ff} & L_{ft} \\ L_{tf} & L_{tt} \end{bmatrix}, \text{with } L_{ab} = \frac{\partial^2 L}{\partial a \partial b} \qquad (2)$$

where L is an image obtained convolving the Logf spectrogram Y with $(3{\times}3)$ derivative masks along frequency f and temporal t directions followed by smoothing using a Gaussian filter with bandwidth σ_D [7]. Here, the L_{ft} represent the second-order derivatives in the f and t directions. Then the contour intensity at each point (f,t) is measured as

$$I_\gamma = (\lambda_2)^{2\gamma}((L_{ff} - L_{tt})^2 + 4L_{ft}^2) \qquad (3)$$

where the eigenvalues of the (2×2) Hessian matrix in Eq. (2) is obtained as

$$\lambda_{1,2} = \frac{1}{2}(L_{ff} + L_{tt}) \pm \frac{1}{2}\sqrt{4L_{ft}^2 + (L_{ff} - L_{tt})^2} \qquad (4)$$

The parameters $\sigma_D{=}$ 10 and $\gamma{=}0.5$ are set, respectively. Here, the optimum N_o is thus automatically selected from $N_o{=}[6, 8, \ldots, 26, 28]$ as the one giving the highest I_γ. Here, the Logf scale oversamples the spectrum with H can be directly specified. The proposed Logf spectrogram thus provides a reconfigurable scheme adapted to the signal as compared to wavelet transform or other constant-Q transform that has intrinsic Logf scaling.

2.2. Local Similarity Measure using Mutual Information

The local similarity measure is performed here by measuring the mutual information (MI) on the Logf spectrogram image. The MI between the two successive image blocks is defined in terms of their joint probability density function (pdf) and the marginal pdf's and is a natural measure of the interdependence or "similarity" between the two image blocks. At block b_t of size $(l{\times}l)$ samples, a matrix $C_{t1,t2}$ can be generated based on gray-scale G-level transitions between the blocks b_{t1} and b_{t2} with G $=256$ here. The element of $C_{t1,t2}$ corresponds to the probability that a pixel with gray level i in block b_{t1} has gray level j in block b_{t2}, with $0 \leq i \leq G-1$ and $0 \leq j \leq G-1$. In other words, $C_{t1,t2}$ is a number of pixels which change from gray level i in block b_{t1} to gray level j in block b_{t2}, divided by the number of pixels in the window block. Following Eq.(5), the MI $A_{t1,t2}$ of the transition from block i to block j is expressed as

$$A_{t1,t2} = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} C_{t1,t2}(i,j)\log\frac{C_{t1,t2}(i,j)}{C_{t1}(i)C_{t2}(j)} \qquad (5)$$

In this paper, the MI values $I_{t_c,t_{c+1}}$ are calculated with the center of block t_c shifted by l samples along time for each non-overlapping frequency bin of width l. It can be noted that the advantage of the MI based similarity measure over the traditional linear methods of analysis like correlation is due to its less sensitivity against outliers as well as to transformations such as scaling, translation [8].

2.3. Feature Selection by 2D-PCA Based Reconstruction

The two-dimensional PCA (2DPCA) computes the eigenvectors of so called image covariance matrix, the size of which is as the same as the width of matrix representation data. This reduces the time and space complexities over PCA and thus improves the feature selection results of data in matrix representation.

Suppose that there are K training images in total, the kth training image is denoted by an $m{\times}n$ matrix $A_k(k = 1, 2, \cdots, K)$, and the average image of all training samples is denoted by $\bar{A}{=}\frac{1}{K}\sum_{k=1}^{K} A_k$. Then, the image covariance matrix, Cov can be estimated as

$$Cov = \frac{1}{K}\sum_{k=1}^{K}(A_k - \bar{A})^T(A_k - \bar{A}) \qquad (6)$$

Then an optimal projection matrix P_{opt}, which is composed by the orthogonal eigenvectors $P_1,\cdots,P_d$ of Cov corresponding to the d largest eigenvalues is obtained as $P_{opt}=[P_1,\cdots,P_d]$. So, the 2DPCA basically learns an optimal matrix P_{opt} from a set of training images reflecting the information of images and then projects an $(m\times n)$ image A onto X, yielding an $(m\times d)$ matrix $R=AP$.

2.4. Signal Quantification by 2D Cepstral Features

In this paper, signal quantification is achieved through the formation of a 2D cepstral feature set based on the new reconstructed image plane R. A discrete time sequence $z[t]$ is first obtained as

$$z[t]=\sum_f R[f,t] \tag{7}$$

where f and t are discrete frequency and time indices, respectively. Then the first Q cepstral coefficients, $c[q]$, $0\leq q\leq(Q-1)$ are calculated given by

$$c[q]=\frac{1}{F}\sum_{f=0}^{F-1}\log\left(J[f]\right)\exp(-j2\pi\frac{f}{F}q) \tag{8}$$

$$J[f]=|\sum_t \hat{z}[t]\exp(j2\pi\frac{f}{F}t)|;\quad 0\leq f\leq F-1 \tag{9}$$

with F is the total number of frequency bins, $\hat{z}[t]$ is the normalized sequence of $z[t]$ with unit norm and zero mean, and $|\cdot|$ refers the absolute value.

A new 2D cepstral feature set is then constructed based on weighting the cepstral sequence by its differential values. Considering $c[q]$ as an integrated sequence, we obtain the weighted cepstral feature $c_w[q]=w[q].c[q]$ where $w[q]$ is the weighting function obtained as second-order differential of $c[q]$, i.e. $w[q]=(1-hB^\tau)^2c[q]$, where $h=1$ and B^τ is a shift-operator with lag τ, i.e. $B^\tau\{c[q]\}=c[q-\tau]$ with $\tau=1$. Here, the cepstral domain is bandpass filtered through $w[q]$ which act as a first-order pre-emphasis product filter with $0.9\leq h\leq 1$ to control the degree of filtering.

3. RESPIRATORY SOUND QUANTIFICATION

3.1. Experimental Dataset

In this study, standard RS recordings extracted from R.A.L.E. datasets [9]. These extracted RS signals were captured at various positions over right/left upper/lower posterior chest using electronic stethoscope from 3 healthy and 12 pathological subjects (9 males/6 females, 11 ± 17 years old), and then been re-sampled at Fs=44.1 kHz. A total of 10 normal RS, 20 CAS, and 17 DAS recordings have been adopted to test the performance of the proposed signal quantification scheme on RS signals captured at various recording sites.

3.2. Choice of data block size for local similarity measure

The optimized data block size of $(l\times l)$ used to calculate the local similarity measure in Section 2.2 is chosen here based on Fisher Ratio [10]. The Fisher Ratio is defined as the ratio of the variance of the feature vectors between classes (σ^2_{inter}) to that within classes (σ^2_{intra}):

$$F_r=\frac{\sigma^2_{inter}}{\sigma^2_{intra}}=\frac{(s.\mu_1-s.\mu_2)^2}{s^T\Sigma_1 s+s^T\Sigma_2 s}=\frac{s.(\mu_1-\mu_2)^2}{s^T(\Sigma_1+\Sigma_2)s} \tag{10}$$

where the direction of the expected line giving the maximum class separation, $s=(\Sigma_1+\Sigma_2)^{-1}(\mu_1-\mu_2)$, and μ_1, μ_2 are the mean values and Σ_1, Σ_2 are the covariances of the feature vectors for the two investigating classes. Here, the variation of the corresponding Fisher Ratio with changing block size on the experimental dataset (i.e. normal and abnormal RS (i.e. DAS or CAS)) has been summarized in Table 1, and block size l=9 which gives the highest Fisher Ratio F_r in both cases has been adopted here.

Table 1. The data block size ($l\times l$) vs. Fisher Ratio, F_r.

Data Types	$F_r(\times 10^4)$				
	l=3	l=5	l=7	l=9	l=11
Normal RS v.s. DAC	0.40	1.01	2.01	2.02	1.30
Normal RS v.s. CAS	6.65	7.05	7.67	8.58	8.29

3.3. Quantification based on Enhanced Cluster Features

The optimized Logf spectrograms of various types of unsegmented RS signals are illustrated in Fig. 2. As compared to STFT spectrogram, enhanced discrimination between normal and abnormal RS signals can be observed in the Logf spectrogram. This shows the effectiveness of the proposed 2D representation and thus enables the extraction of discriminative features based on similarity measures within this 2D image.

The effectiveness of the similarity measure in RS monitoring has been demonstrated in Fig. 3. In order to investigate abnormality detection, the corresponding marginal value of mutual information matrix, $A_f[t]=\sum_f A[f,t]$ has been computed and displayed in Fig. 3(c) with threshold $\overline{A_f}=mean(A_f[t])$. As depicted in the plots, the wheeze episodes could be detected with relatively high accuracy by just simply hard thresholding the marginal similarity value by global mean. The slight displacement of the episodes boundaries are due to the block-wise computation of the mutual information.

Furthermore, the performance of the proposed feature set $c_w[q]$ is shown in Fig. 4 on the experimental dataset. Separability index (SI) [11], being the fraction of data points whose labels are the same as those of their nearest neighbors, is adopted to quantify the performance. With SI=1 indicating completely-separated clusters, it is seen that the unsegmented RS signals can be well separated into two clusters (normal/abnormal) by adopting $c_w[1]$ and $c_w[2]$ only. It can be noted that the weighting function $w[q]$ exploits various differentiating effects on the unweighted $c[q]$ as driven by the variability of data used towards effective quantification.

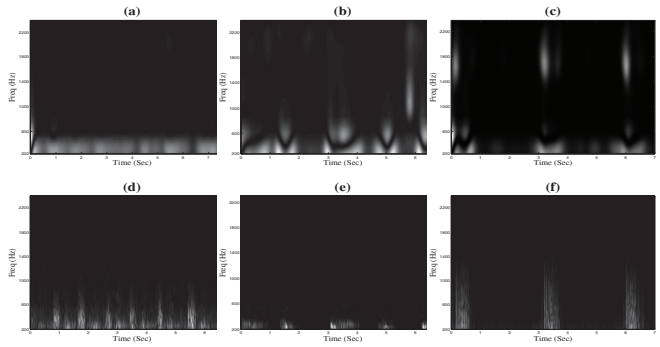

Fig. 2. (Top) The optimized Logf spectrogram of (a) a normal RS signal, (b) a wheeze (CAS) signal (c) a fine crackle (DAS) signal. (Bottom) (d-f) the normal STFT spectrogram of the same set of RS signals.

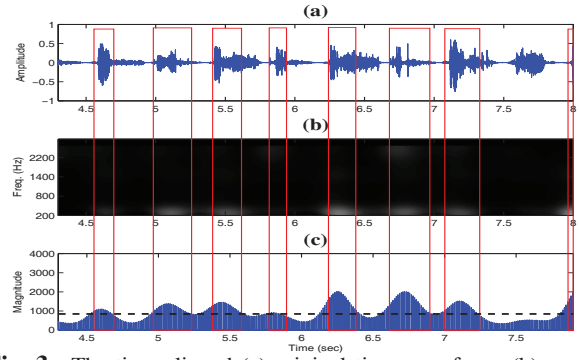

Fig. 3. The time aligned (a) original time waveform, (b) mutual information matrix A, and (c) the corresponding margin value $A_f[t]$ with threshold $\overline{A_f}$ (black dashline) of an unsegmented wheeze signal. The wheeze episodes have been manually detected and labeled with red vertical boxes.

4. CONCLUSION AND FUTURE WORK

An optimized log-frequency spectrogram with signal variability enhanced in TF domain is proposed here.The cepstral plan is expanded into Logf spectrogram with embedded time dimension to better represent the non-stationary audio signals. A new feature selection approach is followed to give discriminative compact feature set that is used for signal quantification. The local similarity measures between pairs of blocks for the optimized Logf spectrogram image acts as a whitening process for the 2D cepstral plane, and therefore allows effective 2D-PCA based reconstruction of the mapped image for feature selection. Experimental results illustrate the low dimensional, highly discriminative cepstral feature set obtained from the reconstructed image is ready for binary clustering toward abnormality detection using unsegmented signals.

Furthermore, since various spectral locations have been associated to various types of RS, we will expand the MI calculation to incorporate geometric information of the pixels and thus improve the robustness of the method and possible differentiation of different types of abnormal RS. Also, longer signals will be considered to test the viability of the method on long-term RS monitoring.

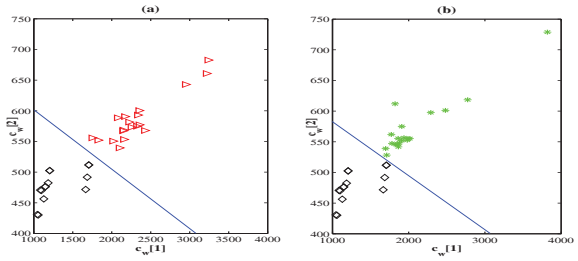

Fig. 4. The scatter plot of (a) Normal RS (black) v.s. CAS signal (red), with SI=1; and (b) Normal RS (black) v.s. DAS Signal (Green) using $c_w[1]$ and $c_w[2]$, with SI=1.

5. REFERENCES

[1] RL Murphy and SK Holford, "Lung sounds," *ATS News*, vol. 8(4), pp. 24–29, 1980.

[2] ARA Sovijarvi, J Vanderschoot, and JR Eavis, "Standardization of computerized respiratory sound analysis," *Eur Resp Rev*, vol. 10(77), pp. 585–649, 2000.

[3] SA Taplidou and LJ Hadjileontiadis, "Analysis of wheezes using wavelet higher order spectral features," *IEEE Trans Biomed Eng*, vol. 57(7), pp. 1596–1610, 2010.

[4] A Homs-Corbera, JA Fiz, J Morera, and R Jané, "Time-frequency detection and analysis of wheezes during forced exhalation," *IEEE Trans Biomed Eng*, vol. 51, no. 1, pp. 182–186, 2005.

[5] SA Taplidou and LJ Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds," *Comput Bio Med*, vol. 37, pp. 1073–1083, 2007.

[6] M Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Comput Biol Med*, vol. 39, pp. 824–843, 2009.

[7] T Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, 1994.

[8] F Maes, A Collignon, D Vandermeulen, G Marchal, and P Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans Med Imag*, vol. 16(2), pp. 187–198, 1997.

[9] PixSoft, *The R.A.L.E. Repository*, http://www.rale.ca.

[10] VN Vapnik, *The Nature of Statistical Learning Theory*, Springer Berlin Heidelberg, New York, 1995.

[11] C Thornton, *Truth from Trash: How Learning Makes Sense*, MIT Press, 2002.