

A multiresolution analysis for detection of abnormal lung sounds

Dimitra Emmanouilidou¹, Kailash Patil¹, James West¹ and Mounya Elhilali¹

Abstract—Automated analysis and detection of abnormal lung sound patterns has great potential for improving access to standardized diagnosis of pulmonary diseases, especially in low-resource settings. In the current study, we develop signal processing tools for analysis of paediatric auscultations recorded under non-ideal noisy conditions. The proposed model is based on a biomimetic multi-resolution analysis of the spectro-temporal modulation details in lung sounds. The methodology provides a detailed description of *joint* spectral and temporal variations in the signal and proves to be more robust than frequency-based techniques in distinguishing crackles and wheezes from normal breathing sounds.

I. INTRODUCTION

Lung and respiratory sounds contain crucial information about pathologies of lungs and/or airway obstruction [1], [2]. Chest examination is the widely used method for diagnosis of pulmonary conditions. However, acoustic information captured by clinical auscultation is limited in a number of ways, including frequency attenuation due to the stethoscope, inter-observer variability, untrained healthcare providers or subjectivity in differentiating subtle sound patterns. Computerized technologies come as a natural non invasive complimentary diagnostic aid. Automated techniques are not only advantageous for providing standardized methods for electronic auscultation, they also enable long duration monitoring and subsequent analysis allowing for a deeper understanding of the mechanisms that produce abnormal pulmonary sounds.

Lung sounds typically span the range of 50 to 2500Hz and consist of multiple components originating from different sources within the respiratory system. Normal breathing sounds generally follow cyclic patterns indicating the airflow during respiratory cycles. In cases of pulmonary diseases, these sounds are often superimposed with anomalous patterns reflecting airway obstructions or pathological conditions. The nature of the anomalous sounds varies between stationary events such as wheezes or rhonchus, to transient sounds such

as crackles, both at fine and coarse scales. Of all lung sounds of interest for medical diagnosis, wheezes and crackles are the most studied components and are indicative of specific pathologies [3], [4]. Wheezes are generally louder than the underlying breathing sound with duration longer than 250ms and are usually signs of obstructive pulmonary diseases (such as asthma). Crackles are transient and explosive sounds, related to the sudden opening of abnormally closed airways and could be indicative of conditions such as pneumonia or lung infection. Their short duration (< 20 ms) and lower intensity introduce difficulties in their discrimination and characterization. Despite their differences, they are not well-defined patterns from a signal processing point of view. Wheezes have been reported to span a wide range of frequencies: 100-2500 or 400-1600Hz [5], [6]. Similarly, crackles have been characterized over a lower but also not well defined range spanning 100-500Hz [5], [7]. Such variability is even harder to characterize in case of paediatric auscultations.

Existing approaches in the literature use techniques to capture the spectral and temporal details of sounds like wheezes and crackles, ranging from frequency analysis using Fourier transform [8], [9], to time-frequency and Wavelet analysis [10], [11], [12]. Other techniques apply image processing methods on the sound spectrograms [13] or compare with reference signals. In most studies however, auscultation recordings often correspond to adults, acquired in a controlled near ideal environment with well defined diagnoses where noise was of minor concern. In the present study, we aim to obtain better insight into the signal characteristics of anomalous lung sounds from pediatric auscultations recorded from infants in Nepal in non ideal conditions, contaminated by crying, background chatter and environmental noise. The study presents an alternative signal processing scheme and develops an analysis methodology for assessment of model accuracy for detecting adventitious sounds and distinguishing them from normal breathing patterns.

II. METHODS

A. Multiresolution Analysis

The framework presented here is based on biomimetic analysis of sound signals believed to take place along the auditory pathway from the point the signal reaches the ear, all the way to central auditory stages up to auditory cortex. Briefly, sound signals $s(t)$ are analyzed through a bank of 128 cochlear filters $h(t; f)$, modeled as constant-Q asymmetric bandpass filters equally spaced on a logarithmic frequency scale spanning 5.3 octaves. The cochlear output is then transduced into inner hair cell potentials via a high and low pass operation. The resulting auditory nerve

*The following are members of the investigator group for this project: James M. Tielsch, Department of International Health, Johns Hopkins Bloomberg School of Public Health; Sudha Basnet, Department of Pediatrics and Child Health, Institute of Medicine, Tribhuvan University, Kathmandu, Nepal; Laxman Shrestha, Department of Pediatrics and Child Health, Institute of Medicine, Tribhuvan University, Kathmandu, Nepal; Subarna K. Khatri, Nepal Nutrition Intervention Project Sarlahi, Kathmandu, Nepal; Joanne Katz, Department of International Health, Johns Hopkins Bloomberg School of Public Health; William Checkley, Division of Pulmonary and Critical Care Medicine, Johns Hopkins School of Medicine; Carolyn Scraftford, Department of International Health, Johns Hopkins Bloomberg School of Public Health. This project was supported by grant number OPP1017682 from the Bill and Melinda Gates Foundation (J. Tielsch, PI); and partial support from NSF CAREER IIS-0846112, AFOSR FA9550-09-1-0234, NIH 1R01AG036424 and ONR N000141010278.

¹Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

signals undergo further spectral sharpening modeled as first-difference between adjacent frequency channels followed by half-wave rectification. Finally, a midbrain model resulting in additional loss in phase locking is performed using short term integration (or low-pass operator $\mu(t; \tau)$ with constant $\tau=2\text{msec}$) resulting in a time frequency representation, the auditory spectrogram (1). More details can be found in [14].

$$y(t, f) = \max[\partial_f(\partial_t(s(t) * h(t; f))), 0] * \mu(t; \tau) \quad (1)$$

At the central auditory stages, cortical neurons analyze details of the spectrographic representation, particularly the signal changes or modulations along both time and frequency. This operation is modeled as 2D affine Wavelet transform. Each filter is tuned ($Q=1$) to a specific temporal modulation ω_0 (or rate in Hz) and spectral modulation Ω_0 (or scale in cycles/octave or c/o), as well as directional orientation in time-frequency space (+ for upward and – for downward). For input spectrogram $y(t; f)$, the response of each cortical neuron is given by:

$$r_{\pm}(t, f; \omega_0, \Omega_0) = y(t, f) *_{t,f} STRF_{\pm}(t, f; \omega_0, \Omega_0) \quad (2)$$

where $*_{t,f}$ corresponds to convolution in time and frequency and $STRF_{\pm}$ is the 2D filter response of each cortical neuron. The resulting cortical representation is a mapping of the sound from a one-dimensional time waveform onto a high-dimensional space. In the current implementation, signals were sampled at 8KHz and parsed onto 3-sec segments. For the two-class problem described later, the model included rate filters covering 0.5-32Hz in logarithmic resolution. For the multi-class problem the cortical analysis included 10 rate filters in the range of 40-256Hz and 7 scale filters in 0.125-8 c/o , also in logarithmic steps. The resulting cortical representation was integrated over time to maintain only three axes of rate-scale-frequency (R-S-F) and was augmented with a nonlinear statistical analysis using support vector machine (SVM) with radial basis function (RBF) kernels [15]. Briefly, SVMs are classifiers that learn to separate the patterns of cortical responses caused by the lung sounds. The use of RBF kernels is a standard technique that allows one to map data from the original space onto a new linearly separable representational space. In the 2-class problem, normal versus abnormal segments were considered. In the multi-class problem categorization was divided into normal, crackle and wheeze sounds where 3 binary classifiers one-versus-all were build, SVM_{ij} , $i, j \in \{1, 2, 3\}$, $i \neq j$. The final decision was based on a majority voting strategy. Each model performance was measured through a 10-fold cross validation with data split into 90-10% for training and testing.

B. Performance Analysis

For the diagnostic accuracy of the model different performance measures were used, all averaged over 10 independent Monte Carlo runs. In all cases the classification rates (CRs) are reported. For the two-class problem in particular, sensitivity (Sens), specificity (Spec) and AUC, the area under the Receiving Operating Characteristic curve (ROC) were used. For the three-class, the 3-way ROC analysis proposed

by Mossman [16] was calculated. Mossman established an analogy of the 2-class ROC analysis to the 3-class case, where the volume under the ROC surface (VUS) expressed the probability that three chosen examples, one each from class 1, 2 and 3, will be classified correctly. Each example is represented by a triplet of probabilities (p_1, p_2, p_3) , where $\sum_i p_i=1$, and $p_i=P(y=i|x)$ expressing the confidence that example x with label y belongs in class i . Plotting these triplets in a three dimensional coordinate space, all examples are bounded by the triangle with triplets (1,0,0), (0,1,0), (0,0,1). These vertices signify a 100% confidence that an example belongs to class 1, 2 or 3 respectively. VUS was obtained using Mossman’s decision rule III on all randomly drawn trios: a trio of examples from each class 1, 2 and 3 is considered correctly rated if the sum of the lengths of the three line segments connecting each triplet with the triangle corner associated to its class is smaller than using any other combination to connect these triplets to the triangle corners. A discriminating test based on chance would obtain $VUS=1/6$. As proposed in [17], to compute p_i , we considered each one of the 2-class SVM_{ij} that discriminates between class i and j , with $i \neq j$, $i, j \in \{1, 2, 3\}$. We first need to find the pairwise class probabilities $p_{ij}=P(y=i|y=i \text{ or } j, x)$, that vector x belongs in class i given SVM_{ij} and x . Assuming that distance d of x from the hyperplane, as outputted by SVM_{ij} is as informative as the input vector x , we estimate these probabilities by $\hat{p}_{ij}=P(y=i|y=i \text{ or } j, d)$, by normalizing the corresponding distances to $[0, 1]$. High probabilities are assigned to examples with greater distances. Notice that $p_{ji}=1-p_{ij}$. Having attained $p_{ij} \approx \hat{p}_{ij}$ for every (i, j) pair, we seek the three posterior probabilities $p_i=P(y=i|x)$. With $k=3$ classes, $P(\bigcup_{j=1}^k y=j|x)=P(\bigcup_{j=1, j \neq i}^k (y=j) \cup (y=i)|x)=1$, and $\sum_{j=1, j \neq i}^k P((y=i) \cup (y=j)|x) - (k-2)P(y=i|x)=1$, which yields: $p_i = \frac{1}{\sum_{j=1, j \neq i}^k (p_{ij})^{-1} - (k-2)}$. All p_i were further normalized so that $\sum_i p_i=1$ holds, and express the confidence about the true class of example x .

C. Dataset

The chest sounds were acquired with a digital recording stethoscope ThinkLabs Inc. connected to a MP3 player at 44,1 KHz sampling rate in the noisy environment of a busy hospital outpatient pediatric clinic in Kathmandu, Nepal. Subjects were young children, healthy or with lower respiratory illness. Two physicians annotated all cases and a total number of 28 recordings of 15-sec duration were selected: 10 normal, 10 wheeze and 8 crackle cases. Our model does not yet include a denoising phase, a necessary step especially considering pediatric auscultations in this non ideal acquisition setup. To correctly evaluate the proposed method cases containing only noise or without signal recording (due to child movement) were excluded. All acquired lung signals were downsampled and split as discussed earlier. Each 3-sec segment maintained the same annotation as the original; notice however that wheezes or crackles mostly occurred during a short period within the sound segment, resulting in

many excerpts being annotated as having an abnormal sound without an actual event being present. Conductive and background noise-talking and/or crying, was strongly apparent, rendering the accurate discrimination between normal and abnormal sounds a difficult task.

III. RESULTS AND DISCUSSION

The joint R-S-F representation of each sound was considered based on the cortical model presented earlier, where the SVM algorithm was able to discriminate normal and abnormal cases with Sens=89.44% and Spec=80.50%. To compare the benefit of the rate-scale representation over existing techniques, the feature extraction method proposed in [9] was applied: the power spectrum of each excerpt was obtained and summed along the frequency axis ranging from 0-800Hz to form a feature vector. The authors used a neural network with two hidden layers for data classification. Since the focus is on the feature parametrization of lung sounds, we need the same SVM backend to compare our method to that from Waitman *et al.*, called Spectral System (SS) in this study. In line with the analysis in [9], we analyzed 3-sec segments with the SS system but varied the feature vector lengths from 10 to 100. Best average performance was achieved for length 90. The AUC values were 0.9217 for the R-S-F and 0.7761 for the SS model. Summary results are presented in Table I in the form of a confusion matrix. Columns correspond to outcomes, rows to true annotations and the diagonal depicts the correct classification % rates.

In order to understand the difficulties of classification of the lung sounds and the ability of the proposed feature dimensions to capture the lung sound characteristics, Fig. 1 is presented. Fig. 1(a) shows a spectrogram of a normal subject. Immediately clear are circular breathing patterns. However, also apparent are noise-like patterns (time 1.2-2.8 sec) that could be easily confused with transient events like crackles. The right panel shows the rate-scale representation based on the cortical analysis of the same signal segment. The figure highlights the presence of a periodic breathing cycle at 4Hz. Strong energy at both positive and negative temporal modulations suggests that the signal fluctuates at 4Hz with no particular upward or downward orientation. Spectrally, the rate-scale pattern shows a concentration of energy in lower scales (<1c/o). This pattern is again reflective of the broadband-like nature of breathing patterns as well as transient noise events. In contrast, Fig. 1(b)-(c) depict similar spectrograms and rate-scale patterns for a diagnosed crackle and wheeze case, respectively. The spectrograms of both cases contain patterns that may easily be confused as wheeze-like. Fig. 1(b) at time 2.5-2.8sec depicts a "crying" interval, not easily discernible as a non wheeze event (contrast with 1.1-1.7sec of Fig. 1(c)). On the other hand, the asymmetry of the rate-scale pattern for both cases begins to show clearer distinction between normal and crackling and wheezing events. Note that the colorbar of rate-scale plots on all cases are different, even though spectrograms were normalized to same level. This is indicative of differences in modulation strength in the signal along both time and frequency.

TABLE I
AVERAGE CLASSIFICATION RATES % OF THE 2-CLASS PROBLEM

True Annotation	R-S-F Output		SS Output	
	Normal	Abnormal	Normal	Abnormal
Normal	80.50	19.50	70.25	29.75
Abnormal	10.56	89.44	18.33	81.67

Average AUC values for R-S-F: 0.9217 and for SS: 0.7761

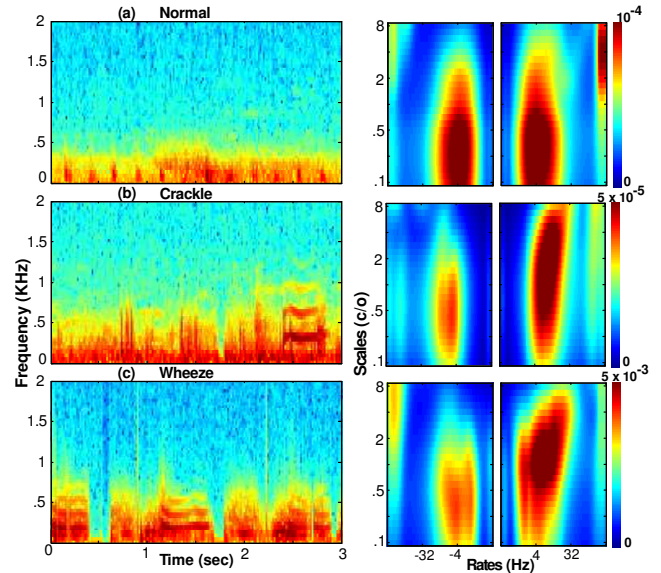


Fig. 1. Time-frequency (column 1) and rate-scale representation (column 2) of data used in the study, showing a normal(a), a crackle(b) and a wheeze(c) case. Notice the difficulties in discrimination based on only time-frequency contents. However distinction is clearer in the rate-scale representation due to the asymmetry noticed comparing the normal with the adventitious cases and a detailed look would indicate stronger components for crackles.

The importance of these feature dimensions can be further investigated considering the more difficult task of a 3-class problem. Exploiting both symmetry and intensity differences in the rate scale representation a discrimination among normal, crackles and wheeze segments is possible through the joint R-S-F setup, as described in the methods, yielding a VUS score 0.601. Recall that a classifier based on chance would achieve VUS=0.167. Detailed CRs may be found in Table II revealing that crackle segments (explosive and short in duration) are more difficult to discriminate, often confused with noise contaminating normal segments. To judge the significance of the frequency components information in the feature vector, all frequency information was averaged resulting in the joint Rate-Scale (R-S) representation (patterns in column 2 of Fig. 1), with VUS=0.729 and CRs as shown in Table II. Corresponding results on the 2-class problem showed Sens=90.22%, Spec=73.50%, AUC=0.9219. A possible reason for the performance jump in the 3-class case compared to the R-S-F representation, could be that knowledge of the specific frequency bands of the abnormal sounds add non-informative details to the model. Another explanation could be the sound set size: not having access to adequate number of abnormal lung sound recordings with enough frequency range variability could be

TABLE II
AVERAGE CLASSIFICATION RATES % OF THE 3-CLASS PROBLEM FOR THE POPOSED METHOD

True Annotation	R-S-F Output			R-S Output			[R,S] Output		
	Normal	Crackle	Wheeze	Normal	Crackle	Wheeze	Normal	Crackle	Wheeze
Normal	77.49	10.01	12.50	76.75	16.50	6.75	76.07	11.79	12.14
Crackle	21.25	39.25	39.50	19.50	45.76	34.74	20.71	50.35	28.93
Wheeze	20.60	23.20	56.20	10.02	16.38	73.60	10.00	21.71	68.29

Average VUS values for R-S-F: 0.601, for R-S: 0.729 and for [R,S]: 0.608

also affecting the performance. It is unclear at this point how much frequency localized are the specific sound patterns and whether the frequency coverage correlates with specific pathological or ecological substrates. Further investigation is ongoing to gain more insight into the nature of the data.

Finally, we closely investigated the importance of having such a joint spectro-temporal modulation space. To this effect, we assess the relevance of the marginal feature dimensions for rates and scales, where we consider the rate alone feature vector extending the scale representation ([R,S]) and the achieved 3-class VUS score was 0.6075 and the 2-class AUC was 0.8572. CRs are shown in Table II. We note that the joint R-S representation appears more informative in discriminating between the sounds of interest, compared to both rates and scales in a concatenated vector.

IV. CONCLUSION

An automated multi resolution analysis of lung sounds was introduced in this study. While the majority of the literature methods is based on extracting frequency features and analyzing spectrograms or other time frequency representations, this study proposes a spectro-temporal modulation feature extraction, inspired from auditory cortical representations. A real life application is considered with clinical pediatric auscultation performed in non controlled noisy environments, capturing possible variations of sound events in spontaneous breathing conditions. SVM classifiers were trained on the different extracted features and evaluated using correct classification rates and VUS scores, measuring the discriminating ability among normal, crackle and wheeze cases.

The observed results revealed that lung sounds contain more informative details than the time-frequency domain can capture. Temporal and spectral modulation features are able to increase the discrimination capability compared to features based only on frequency axis. A joint rate-scale representation is able to perform sufficient discrimination even in noisy sound segments where talking and crying can impede or complicate the identification of abnormal sounds. This is the first step to a deeper understanding of the signal processing characteristics defining sound events in lung recordings, yielding promising results. Continued focus will be to pre-process sounds applying denoising techniques, specific oriented to background and conductive noise. Further work will be to extract the breathing cycle and isolate events related to inspiration or expiration, leading to a more accurate discrimination of abnormal sounds, as indicators of pulmonary conditions and their severity.

REFERENCES

- [1] A. R. A. Sovijärvi, L. P. Malmberg, G. Charbonneau, and J. Vanderschoot, "Characteristics of breath sounds and adventitious respiratory sounds," pp. 591–596, 2000.
- [2] L. E. Ellington, R. H. Gilman, J. M. Tielsch *et al.*, "Computerised lung sound analysis to improve the specificity of paediatric pneumonia diagnosis in resource-poor settings: protocol and methods for an observational study," *BMJ open*, vol. 2, p. e000506, 2012.
- [3] A. A. Abaza, J. B. Day, J. S. Reynolds, A. M. Mahmoud, W. T. Goldsmith, W. G. McKinney, E. L. Petsonk, and D. G. Frazer, "Classification of voluntary cough sound and airflow patterns for detecting abnormal pulmonary function," *Cough*, vol. 5, p. 8, 2009.
- [4] A. Gurung, C. G. Scrafford, J. M. Tielsch, O. S. Levine, and W. Checkley, "Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis," *Respir Med*, vol. 105, no. 9, pp. 1396–1403, Sep 2011.
- [5] S. Reichert, R. Gass, C. Brandt, and E. Andrés, "Analysis of respiratory sounds: state of the art," *Clinical medicine Circulatory respiratory and pulmonary medicine*, vol. 2, pp. 45–58, 2008.
- [6] S. A. Taplidou, L. J. Hadjileontiadis, I. K. Kitsas *et al.*, "On applying continuous wavelet transform in wheeze analysis," *Conference Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society*, vol. 5, pp. 3832–3835, 2004.
- [7] B. Flietstra, N. Markuzon, A. Vyshedskiy, and R. Murphy, "Automated analysis of crackles in patients with interstitial pulmonary fibrosis," *Pulmonary medicine*, vol. 2011, no. 2, p. 590506.
- [8] K. K. Guntupalli, P. M. Alapat, V. D. Bandi, and I. Kushnir, "Validation of automatic wheeze detection in patients with obstructed airways and in healthy subjects," *The Journal of asthma official journal of the Association for the Care of Asthma*, vol. 45, pp. 903–907, 2008.
- [9] L. R. Waitman, K. P. Clarkson, J. A. Barwise, and P. H. King, "Representation and classification of breath sounds recorded in an intensive care setting using neural networks," *Journal of Clinical Monitoring and Computing*, vol. 16, no. 2, pp. 95–105, 2000.
- [10] Y. P. Kahya, M. Yeginer, and B. Bilgic, "Classifying respiratory sounds with different feature sets," *Conference Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 2856–2859, 2006.
- [11] X. Lu and M. Bahoura, "An integrated automated system for crackles extraction and classification," *Biomedical Signal Processing And Control*, vol. 3, no. 3, pp. 244–254, 2008.
- [12] A. Kandaswamy, C. S. C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients," *Computers in Biology and Medicine*, vol. 34, no. 6, pp. 523–537, 2004.
- [13] R. J. Riella, P. Nohama, and J. M. Maia, "Method for automatic detection of wheezing in lung sounds," *Brazilian journal of medical and biological research Revista brasileira de pesquisas medicas e biologicas Sociedade Brasileira de Biofisica et al*, vol. 42, no. 7, pp. 674–684, 2009.
- [14] T. Chi, P. Ru, and S. Shamma, "Multiresolution spectrotemporal analysis of complex sounds," *J. Acoust. Soc. Am.*, vol. 118, pp. 887–906, 2005.
- [15] N. Cristianini and J. Shawe-Taylor, *Introduction to support vector machines and other kernel-based learning methods*. Cambridge, UK: Cambridge University Press, 2000.
- [16] D. Mossman, "Three-way rocs," *Medical Decision Making*, vol. 19, no. 1, pp. 78–89, 1999.
- [17] D. Price and S. Knerr, "Pairwise neural network classifiers with probabilistic outputs," *Processing*, vol. 7, pp. 1109–1116, 1995.