35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Characterization of Noise Contaminations in Lung Sound Recordings*

Dimitra Emmanouilidou[1] and Mounya Elhilali[1]

Abstract—**Lung sound auscultation in non-ideal or busy clinical settings is challenged by contaminations of environmental noise. Digital pulmonary measurements are inevitably degraded, impeding the physician's work or any further processing of the acquired signals. The task is even harder when the patient population includes young children. Agitation and/or crying are captured into the recordings, additionally to any existing ambient noise. This study focuses on characterizing the different types of signal contaminations, expected to be encountered during lung sound measurements in non-ideal environments. Different noise types were considered, including background talk, radio playing, subject's crying, electronic interference sounds and stethoscope displacement artifacts. The individual characteristics were extracted, discussed and further compared to characteristics of clean segments. Additional exploration of discriminatory features led to a spectro-temporal signal representation followed by a standard SVM classifier. Although pulmonary and ambient sounds were both dominant in most sound clips, such a complex representation was deemed to be adequate, capturing most of the signal's distinguishing characteristics.**

I. Introduction

Lung sound auscultation has been a valuable part of clinical assessment for patients. It is usually the first tool used by primary care providers as it can reveal lung diseases in a noninvasive and cost-effective manner simply by listening to the chest sounds. Respiratory and lung diseases are a major public health concern in both industrial and developing countries, though the latter usually lacks experienced or well-trained clinical personnel. The challenge in such settings is the high inter observer variability in interpreting sound content as captured by the stethoscope, as well as the many different sources of noise contamination. In contrast to well-controlled clinical environments where noise is of little or no concern, when auscultation is performed in outpatient or busy clinics, the signal can be significantly corrupted or degraded by environmental sounds, thus impeding the work of the physician. In addition, when pediatric auscultation is considered, agitation, movement and cry can be most prominent throughout auscultation.

Computer aided analysis offers the advantages of meticulous, offline revision and further processing of the recorded signal, towards noise reduction and identification of events-indicators of possible pulmonary disease or dysfunction. A lot of work has been published on lung sound signal denoising, but mostly focused on reducing the heart sounds or identifying adventitious events. To the best of our knowledge, limited literature has been found to address pediatric auscultation in non-ideal settings. Bahoura etal. [1] proposed a denoising technique using Wavelet Packets on white and instrumentation/ventilation noise; Suzuki etal. implemented an adaptive filter with the use of a reference recording, applied on an adult recording exposed in background radio talking [2]. In order to better understand the nature of these potential contaminations, the current study focuses on characterizing different types of noise being captured during digital auscultation, when subjects are young children and data are acquired in busy non-ideal environments. Signal contaminations considered here involve ambient noise, background talking, crying, electronic interference and artifacts produced by intentional or unintentional stethoscope displacements.

* 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.

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

II. Methods

Data were obtained from a pool of lung sound recordings acquired in a children's hospital in Lima, Peru. More information on the acquisition protocols can be found in [3]. 53 subjects (control cases) were considered in the current study. A digital recording stethoscope of ThinkLabs Inc. connected to an MP3 player at 44.1 KHz sampling rate was used for the acquisition. All sounds were then downsampled to 8 KHz. Short sound segments with duration of 0.5-3 sec were manually extracted from various recording segments within the signal, including left/right anterior/posterior inferior/superior sites. Samples, consisting of noise- and lung sound-related content, where the latter contained no kind of adventitious events, were divided into 5 categories. The first one, $Clean_B$, included 'clean' lung sound signal. These segments were picked from control patient cases with limited background or other noise. Four further groups were formed to capture distinct sources of signal corruption: $Background_N$, representing any background noise such as background talking, distant children crying, radio playing or children toys' sounds; Cry_N including intervals of crying coming from the child under examination; $Interference_N$, with sound segments contaminated by mobile or other source of electronic interference (buzzing) and finally $StethMove_N$, a group capturing intentional displacement of the stethoscope during the recording, i.e. when the physician changed location of recording site, or unintentional displacement, e.g. when subject appeared to be agitated. Note that $StethMove_N$ group contained limited lung sounds contents which were very prominent in all other categories. All isolated segments were processed into short 500ms-windows with 50% overlap.

A. Spectrum Characterization

The short-time 2^{14}-point Fast Fourier Transform (FFT) was calculated for each sound segment, smoothed with a 5th order Butterworth filter with cutoff frequency at 60 Hz and averaged over all windows. From the smoothed amplitude spectrum, a number of features were extracted:

- Peak Width (PW). The maximum spectrum peak was extracted and its width measured at 75% of its corresponding height. To avoid confusion with high-frequency peaks (see profile of Cry_N in Fig. 1), the search for the maximum peak was restrained to frequencies below 200 Hz.
- Spectrum Slope (SL_{100}). It has been previously shown [4] that the spectrum produced by lung sound recordings decays exponentially with frequencies higher than 75 Hz. These findings came from adult recordings with controlled environmental noise. In our case this threshold was found to be closer to 100 Hz and so it was increased accordingly. The spectrum P, expressed in logarithmic scale as $20 \cdot log\ (P/P_{thr})$, with $P_{thr} = 5 \cdot 10^{-5}$, was fit with a linear regression line and its slope calculated in dB/octave.
- Power Ratio (PR), calculated as the total estimated power versus the power of the regression line. The estimated power at frequency f was expressed as $P_{est}\ (f) = P_{thr} \cdot (f / f_{max})^{SL}$, with f_{max} the point where the logarithmic spectrum curve crosses the frequency axis, as proposed in [4]. The power of the regression line depicts the area underneath the linear regression line described above.
- Low-to-High Frequency Ratio ($LHFR_{500}$), the ratio of average squared power spectrum for frequencies below 500 Hz versus the average power at frequencies above 500 Hz. Lung sound content containing no adventitious events has been found to be concentrated at low frequencies, and thus, this metric was expected to capture frequency content not related to any respiratory or heart sounds [5,6].

B. Harmonicity

In a spectrum amplitude representation of a signal, when spectral components are found at integer multiples of a common low frequency- the fundamental frequency, F_0- they are said to be harmonically related and provide evidence of the harmonic profile of the sound excerpt. In complex sounds like the ones used in this study, possible harmonics are expected at roughly- not necessarily exact- integer multiples of a F_0. The following algorithm was used to capture harmonicity of short term bursts of high energy content: In step 1, the transient events with broadband energy were identified as follows. The short-time Fourier transform of the signal was calculated using 50ms windows with 50% overlap. The spectrum of each segment was then averaged across frequencies above 1 KHz. This cutoff was chosen to exclude most of lung sound-specific information. All instances with non-negligible power were then isolated from the resulting time series, revealing locations of high frequency content. In step 2, a 50ms window centered at each time-peak location was extracted from the original sound waveform, and its 2^9-point FFT was computed. From the calculated spectrum, a sequence of at most 8 peaks was identified, excluding the very first spectrum peak. If at least 80% of the spectral peaks formed a harmonic stack with 20 Hz tolerance, then the time clip was considered to be harmonic. This process was repeated for all time-peak locations of step 1.

C. Spectral and Temporal Modulations

Inspired by recent psychophysical and physiological findings on the way the brain processes sound information travelling from the inner ear all the way to the auditory cortex, a multi-resolution analysis was invoked [7]. The sound signal, $s(t)$, was processed through a bank of 128 overlapping constant Q band-pass filters. The filters, asymmetric with central frequencies uniformly distributed in logarithmic scale covering 5.3 octaves, resemble the processing done in the basilar membrane. After a high and low pass operator mimicking the hair cell stage, the auditory nerve output was spectrally sharpened and integrated over short windows, resulting in a time-frequency representation $y(t, f)$, which is effectively a spectrogram-like representation of the input sound signal $s(t)$. The next stage, representing the higher central processing in the auditory cortex, is mathematically expressed by a 2D affine Wavelet Transform of the spectrogram $y(t, f)$. The spectral and temporal modulations of the auditor spectrogram were calculated using a bank of Gabor like modulation-selective wavelet filters. Each of these filters was tuned to a specific temporal (rates, ω, in Hz) and spectral (scales, Ω, in cycles/octave or c/o) modulation. A bank of directional (±) selective filters was used to capture content changing in positive or negative phase, with a corresponding spectro-temporal impulse response $STRF_{\pm}(t, f, \omega, \Omega)$. The final spectro-temporal representation forms the cortical response of the input spectrogram $y(t, f)$ and was calculated as:

$$cr_{\pm}(t, f, \omega, \Omega) = y(t, f) *_{t,f} STRF_{\pm}(t, f, \omega, \Omega) \quad (1)$$

using $*_{t,f}$ to denote double convolution in time and frequency. Throughout the paper, the magnitude of the cortical response has been used. The time axis was then integrated over windows of 500 ms, yielding a scale-rate-frequency representation, S-R-F, which is then averaged across all windows:

$$srf_{\pm}(\omega, \Omega, f) = \sum_n \sum_\tau cr_{\pm}(\tau, f, \omega, \Omega) / (N_n \cdot N_\tau), \quad (2)$$

where $\sum_\tau$ denotes integration over each short time window of size N_τ, and $\sum_n$ integration over all N_n windows. The modulation selective filters were created using 31 distinct rates ω, and 31 scales Ω, all equally spaced in logarithmic axis, with $\omega \in [4, 256]$ Hz; $\Omega \in [0.125, 8]$ c/o.

D. Classification

The supervised learning algorithm Support Vector Machines (SVM), with a radial basis function (RBF) kernel, was invoked [8] to capture the differences among types of noise using the S-R-F representation. Data, transformed by the kernel function to a linearly separable feature space, were split into training and testing examples. During the training

phase a hyperplane margin is constructed to separate examples of the two groups; during the testing phase, data were classified according to their distances from the hyperplane. For our multiclass problem, 10 binary $SVM_{i,j}$ were constructed and trained using examples from every pair of groups i, j, where $i \neq j$ and $SVM_{j,i} \equiv SVM_{i,j}$. A test example k was tested on all $SVM_{i,k}$, $i \neq k$ and classified according to majority vote. Bias was removed by randomizing selection in the event of a tie. To reduce the high dimensional feature space, tensor Singular Value Decomposition (SVD) was applied before classification. Data were unfolded along each feature dimension and the principal components were calculated from the covariance matrix. Components capturing no less than 99% of the total variance were kept to form the reduced feature space dimensions. The dimensionality was therefore reduced from 31x62x128 to at most 6x1.

III. Results

A. Signal Characteristics

Twenty reference samples were considered for each one of the five sound categories listed in section II. Segments were processed into short time windows, as discussed earlier, to extract the individual spectrum characteristics. The mean spectrum profile of each class is shown in Fig. 1(left) and representative examples and spectrum slope plots in Fig. 1 (right). Mean feature values for each group are reported in Table I. Samples of group Cry_N showed a high peak width, PW, and a significant content concentration in higher frequencies, achieving a very small $LHFR_{500}$. Spectrum slope value, SL_{100}, was not very informative in this group since crying profiles were far from being exponentially decaying with frequencies above 100 Hz. The latter was also depicted in the high PR value. Cases of the $StethMove_N$ group showed a steep spectrum slope with increased power ratio when compared to $Clean_B$ or $Interference_N$ groups. The $Clean_B$ group yielded the lower PW, SL_{100}, PR values, with spectrum content mostly concentrated below 500 Hz. As expected, the $Background_N$ group being heavily contaminated with talking and crying revealed increased $LHFR_{500}$ compared to group $Clean_B$, where most spectrum contents were pulmonary-related and in lower frequencies.

B. Harmonic Profile

The spectral features presented provided general evidence of the peculiarities of the distinct noise types. A more detailed look into the profiles of $Interference_N$ noise and $StethMove_N$ artifacts showed isolated or repeated short-time bursts of broadband energy. Listening to these burst of energy in samples of electronic interference a certain 'musicality' emerged, an attribute of signal's harmonicity. Such a characteristic is neither heard nor expected for sounds in the $StethMove_N$ group. The reader is referred to Fig. 1(d) where arrows indicate the evident harmonic profile of a $StethMove_N$ case. The detection algorithm presented in section II.B was applied to first identify transient events of the time-frequency representation and then decide if a harmonic structure was exhibited. Considering the detected harmonic segments of group $Interference_N$ from all case files, a consistent fundamental frequency was found at 215.09Hz (±2.76Hz) after rejecting 10% of possible extreme outliers. There was no obvious harmonic structure observed for the cases of group $StethMove_N$. Figure 2 shows the spectrogram of two case examples, where the identified bursts of energy were marked within black margin regions. Clips exhibiting a harmonic structure are shown with an 'X'.

TABLE I. AVERAGE SPECTRUM FEATURES PER SOUND GROUP

Group	Spectrum Features *			
	PW	***SL_{100}***	***PR***	***$LHFR_{500}$***
$Clean_B$	132.12	-9.14	100.91	27857.58
$Background_N$	162.72	-10.06	407.85	14702.98
Cry_N	209.28	-10.06	286.62	2581.76
$Interference_N$	163.70	-9.76	113.03	9140.24
$StethMove_N$	116.63	-11.78	406.83	8254.19

*PW: peak width, SL_{100}: spectrum slope, PR: power ratio, $LHFR_{500}$: low-to-high frequency ratio.

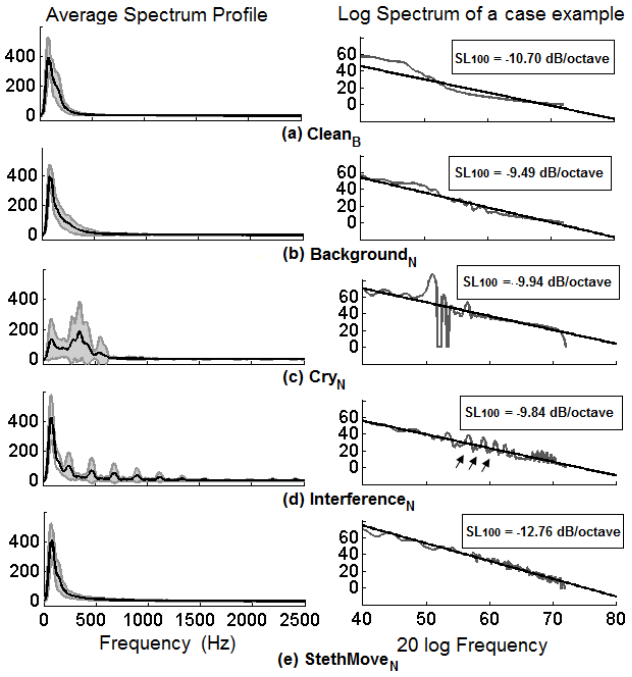

Figure 1. Left panels: average spectrum profile of all sound group. Shaded regions reflect the standard deviation among group cases. Right panels: logarithmic spectrum plot of selected case examples. The dark line represents the linear line fit to the spectrum and the slope shown in legend.

C. Classification of Noise Signals

Spectral and harmonicity features discussed revealed distinct characteristics of the sound samples belonging to each category. Further inquiry on whether the peculiarities of the sound signals could adequately distinguish between the different types of noise led to an SVM classifier. Since most groups share common lung-related content, and sounds belonging in $Clean_B$ and Cry_N groups share a lot with $Background_N$ group, the classification task was expected to be non-trivial. The spectral and temporal modulations of each segment were captured using the reduced S-R-F data representation described in section II.C. Examples were

tested through all respective binary SVMs and the label decision was based on majority vote. Average results and standard deviation over 10 independent runs of a 5-fold cross validation, (80% training data and 20% test data) are shown in Table II. The results depicted the inherited difficulty in discrimination. See for example the confusion in columns $Clean_B$ and $Background_N$ since background noise and lung sound content were apparent in most sound clips. Note that samples in $Interference_N$ group were in majority overwhelmed by background noises containing a weaker interference component. Also, both $Interference_N$ and $StethMove_N$ samples were characterized by transient bursts of broadband energy. Those facts added extra confusion to the classification scheme, and were also depicted in the table of results.

IV. Conclusion

A number of noise factors such as crying, talking, background radio playing, patients movement etc., are rarely or never considered in adult auscultation and well controlled clinic environments, on which the majority of published work relies. However, pediatric auscultation performed in busy environments is inevitably challenged by all the aforementioned factors and it was the purpose of this paper to present, describe and analyze the different signal contaminations expected to be encountered in such settings. A number of feature characteristics were extracted and revealed distinguished patterns for the different noise categories. Although signals from all five sound groups shared a lot of common information, i.e. the actual lung sound content and the background or environmental noise, the features presented above, such as the spectral width, the content concentration within frequency bands, a possible harmonic structure, revealed distinct spectrum characteristics for each specified group. An augmented spectro-temporal representation further supported the statement that noise contaminations encountered in such recordings have distinct features and can be discriminated. For example a strong harmonic profile can reveal probable interference noise; or high energy contents within the range of (200-600) Hz can suggest significant cry contaminations and so on. Incorporating knowledge of all those noise features into computer aided diagnostic tools could contribute to better discrimination between adventitious events and noise contaminations, thus, leading to improved and more robust automated signal analysis and processing techniques.

Acknowledgment

The authors would like to acknowledge the significant contribution of Laura E. Ellington, William Checkley and the rest of the team in Division of Pulmonary and Critical Care, Johns Hopkins University, Baltimore, Maryland and Instituto Nacional de Salud del Niño, Lima, Peru [3] for providing the data and continuous feedback and discussions.

References

[1] M.Bahoura, M Hubin, and M Ketata. "Respiratory sounds denoising using wavelet packets." *Proceedings of the 2nd International Conference on Bioelectromagnetism Cat No98TH8269,* 1998.

[2] A. Suzuki, C. Sumi, K. Nakayama, M. Mori, "Real-time adaptive cancelling of ambient noise in lung sound measurement", Med. Biol. Eng. Comput., vol. 33, no. 5, pp. 704–708, 1995.

[3] 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 anobservational study," BMJ open, vol. 2, p. e000506, 2012.

[4] N. Gavriely, Nissan, M., Rubin, A.H. and Cugell, D.W. Spectral characteristics of chest wall breath sounds in normal subjects, *Thorax* 50, 1292-1300 (1995).

[5] A.R.A. Sovijärvi, L.P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, J.E. Earis, "Characteristics of breath sounds and adventitious respiratory sounds," Eur Respir Rev, vol. 10, 591-6, 2000.

[6] A. Davignon, P. Rautaharju, E. Boisselle, M. Megdlas, and A. Choquette, "Normal ECG Standards for Infants and Children," *Pediatric Cardiology*, vol. 1, no. 2, pp. 123–131, 1979.

[7] T. Chi, P. Ru, and S. Shamma, "Multiresolution spectrotemporal analysis of complex sounds," *Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 887–906, 2005.

[8] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, vol. 1, no. 1. Cambridge University Press, 2000, p. 204.

TABLE II. Average (±STD) Classification Results

True Label	Output				
	$Clean_B$	$Background_N$	Cry_N	$Interference_N$	$Steth Move_N$
$Clean_B$	**95.00 (±5.77)**	2.00 (±2.58)	0	2.00 (±3.50)	1.00 (±3.16)
$Background_N$	3.00 (±3.50)	**91.00 (±7.75)**	0	3.50 (±5.30)	2.50 (±2.64)
Cry_N	1.50 (±3.37)	2.00 (±3.50)	**93.50 (±5.30)**	1.50 (±2.42)	1.50 (±3.37)
$Interference_N$	2.50 (±2.64)	5.50 (±5.50)	2.00 (±3.50)	**85.50 (±6.85)**	4.50 (±4.97)
$Steth Move_N$	3.50 (±4.12)	4.00 (±4.59)	0.50 (±1.58)	1.00 (±2.11)	**91.00 (±6.58)**

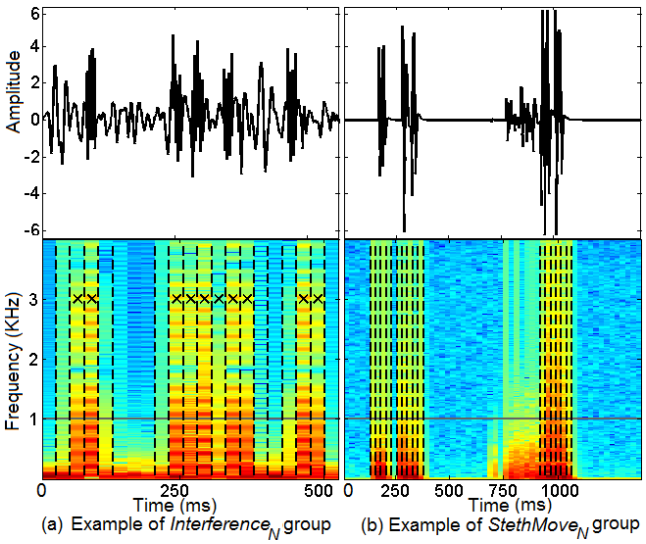

Figure 2. Selected case examples of $Interference_N$(a) and $StethMove_N$(b) groups. The time waveforms (top panels) and corresponding spectrograms (bottom panels) are shown. Black dashed lines mark the identified transient events of broadband energy. Segments found to exhibit a harmonic structure are noted with a 'X'.