IMPROVING THE QUALITY, ANALYSIS AND INTERPRETATION OF BODY SOUNDS ACQUIRED IN CHALLENGING CLINICAL SETTINGS

by

Dimitra Emmanouilidou

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

January, 2018

Abstract

Despite advances in medicine and technology, Acute Lower Respiratory Diseases are a leading cause of sickness and mortality worldwide, highly affecting countries where access to appropriate medical technology and expertise is scarce. Chest auscultation provides a low-cost, non-invasive, widely available tool for the examination of pulmonary health. Despite universal adoption, its use is riddled by a number of issues including subjectivity in interpretation and vulnerability to ambient noise, limiting its diagnostic capability. Digital auscultation and computerized methods come as a natural aid towards overcoming such imposed limitations.

Focused on the challenges, we address the demanding real-life scenario of pediatric lung auscultation in busy clinical settings. Two major objectives lead to our contributions: 1) Can we improve the quality of the delicate auscultated sounds and reduce unwanted noise contamination; 2) Can we augment the screening capabilities of current stethoscopes using computerized lung sound analysis to capture the presence of abnormal breaths, and can we standardize findings. To address the first objective, we developed an adaptive noise suppression scheme that tackles con-

tamination coming from a variety of sources, including subject-centric and electronic artifacts, and environmental noise. The proposed method was validated using objective and subjective measures including an expert reviewer panel and objective signal quality metrics. Results revealed the ability and superiority of the proposed method to i) suppress unwanted noise when compared to state-of-the-art technology, and ii) faithfully maintain the signature of the delicate body sounds.

The second objective was addressed by exploring appropriate feature representations that capture distinct characteristics of body sounds. A biomimetic approach was employed, and the acoustic signal was projected onto high-dimensional spaces spanning time, frequency, temporal dynamics and spectral modulations. Trained classifiers produced localized decisions on these breath content features, indicating lung diseases. Unlike existing literature, our proposed scheme is further able to combine and integrate the localized decisions into individual, patient-level evaluation. A large corpus of annotated patient data was used to validate our approach, demonstrating the superiority of the proposed features and patient evaluation scheme. Overall findings indicate that improved accessible auscultation care is possible, towards creating affordable health care solutions with worldwide impact.

Primary Reader: Dr. Mounya Elhilali

Secondary Reader: Dr. James E. West

Acknowledgments

Work in Chapter 2 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. Work in Chapter 3 was supported by grant OPP1084309 and 48968 from the Bill & Melinda Gates Foundation; grants NSF CAREER IIS-0846112; NIH 1R01AG036424; and ONR N000141010278 and N000141210740. I would like to acknowledge the significant contribution of Eric D McCollum, Laura E. Ellington, William Checkley and the rest of the team in Division of Pulmonary and Critical Care, Bloomberg School of Public Health and School of Medicine, Johns Hopkins University, Baltimore, Maryland, and Instituto Nacional de Salud del Nio, Lima, Peru for providing the data and continuous feedback and discussions; and also Daniel Park, Laura Hammitt, Katherine O Brien, Niranjan Bhat, Stavros Papadopoulos, the PERCH Digital Auscultation Study Group, the PERCH Study Group for the valuable contributions to this work.

My continuing appreciation also goes to all the mentors I have had during my

academic career: Dr. Mounya Elhilali, Dr. James E. West, Dr. Panagiotis Tsakalides, Franco Chiarugi. I also would like to thank all influential people and friends in my life, Ellington West, Arthur T. Ward, Jai McLane, Dan McLane, Lucy Melkonian, Jerard Tourgoutian, Nikos Giannakis, Dimitris Milioris, Aris Efmorfoutsikos, Elina Spyrou, Elina Akalestou, Elena Chrysostomou, Yannis Grammatikakis, Thomas Karathanos, Daniela Rose, Christos Sapsanis, Bill Papaioannou, Haralampos Avraam, Maritina Iliadi, George Gemenetzis, Anahita Mehta, Pegky Foteinou, Spyros Stamatelos, Ani Hristoskova.

Many thanks to my great lab mates, Dr. Michael Carlin, Dr. Merve Kaya, Debmalya Chakrabarty, Ashwin Bellur, Nick Huang, Benjamin Davis, Harshavardhan Sundar, David Little. A big THANKS also to my colleague Ian McLane, a brilliant student and coworker. Thank you for the many late-hour brainstorming sessions, and for making the long days in the lab bearable with your excitement! To my most influential mentor and advisor, words are not enough to express my gratitude for your continuous support and guidance but I will try: Dear Mounya, you have been the best mentor I could ever hope for, since day one. Im proud and grateful to have been part of the LCAP family, and have grown into the person I am today. Your support, your diligence, your guidance, your mentoring, your fairness, your continuing advocacy for your students and your will to see us succeed have enabled me to become a better person, better collaborator, better researcher, better scientist. And most importantly.. I have learnt and promise to always keep my slide animations

to a minimum."

Thank you to the amazing staff of Electrical & Computer Engineering in Johns Hopkins, for advocating for the work and the students of this wonderful institution. Thanks to Dr. R.J. Vogelstein for the wonderful latex template.

Finally, this journey would not have been possible without the unconditional support and love of my family. Thank you mama Popi, papa Steve, big brother Manos, twin brother Nikos, I hope I will continue to make you all proud. Also, thank *you* Robert Gordon for the continuing support and love Ive received from you and all your family; youve somehow managed to make a huge foreign country feel like home.

Contents

List of Tables

List of Figures

Chapter 1

Motivation and Background Information

The use of chest auscultation to diagnose lung infections has been in practice since the invention of the stethoscope in the early 1800s. It is a diagnostic instrument widely used by clinicians to "listen" to lung sounds and flag abnormal patterns that emanate from pathological effects on the lungs. While often complemented by other clinical tools such as chest radiography or other imaging techniques, as well as chest percussion and palpation, the stethoscope remains a key diagnostic device due to its portability, low-cost and its non-invasive nature. Chest auscultation with standard acoustic stethoscopes is not limited to resource-rich industrialized settings. In low-resource high-mortality countries with weak health care systems there is limited access to diagnostic tools like chest radiographs or basic laboratories. As a result, health care providers with variable training and supervision rely upon low-cost clinical tools like standard acoustic stethoscopes to make critical patient management decisions. Despite its universal adoption, the use of the stethoscope is riddled by a number of issues including subjectivity in interpretation of chest sounds, inter-listener variability and inconsistency, need for medical expertise, as well as vulnerability to ambient noise which can

mask the presence of sound patterns of interest.

1.1 Understanding the True Need

Most modern clinics are equipped with advanced facilities and tools of increased sensitivity for pulmonary diagnostics: chest X-Rays can reveal lung abnormalities with increased or decreased density, including consolidation, interstitial or chronic lung diseases; Ultrasounds effectively reveal the presence of fluid excess in the lungs, and help diagnose, among others, consolidation and pneumothorax cases; CT scans can provide high resolution images for assessing lung disorders like COPD, cancer, pneumonia. In places where this technology is accessible, the stethoscope is of little use and one might be tempted to assume that its value is declining. But the stethoscope remains the most used diagnostic tool in a vast number of settings where fast, low-cost and portable solutions are a priority; i) mobile clinics or emergency units, including ambulance vehicles, helicopters, airplanes, space crafts or military units; ii) developing or remote countries with limited access to medical experts or advanced technology. Clinical examination in these settings is hindered by severe ambient noise leaks into the audible body sounds, masking their clinical value, and challenging their medical interpretation. Computerized solutions are of paramount importance for delivering an enhanced, high-quality signal, and for providing robust automated diagnostic assistance. And as the need for computerized systems and automated patient diagnosis has clearly been portrayed in real life scenarios, why havent they taken over the health care system yet? Is medical research still riddled by the challenges of automated auscultation diagnostics, or is there simply a research gap between actual need and available solutions? A closer look into the available literature will help answer this question.

1.2 Review on Previous Work

Literature

Computerized analysis for medical sound interpretation has been blooming over the last few decades; research breakthroughs come from scientific teams all around the globe, and yet, this work has not been sufficient to fully address the requirements of an automated diagnostic aid-tool, whose value lies in its ability to adapt to unseen auscultation protocols and clinical settings, and its requirement of limited or no supervision.

The need for improved auscultation solutions has always been urgent: a study in 1996 assessed lung sound interpretation scores of emergency physicians, as compared to paramedics,[3] revealing that paramedics scored significantly worse in lung sound diagnostics. A year later, another study revealed the same urgent need, this time under challenging settings: physicians breath sound assessment-accuracy dropped by 42% in a moving ambulance when compared to a quiet room assessment.[4] Unfortunately up to date, there is limited work addressing such real-life scenarios: most recent studies consider computerized solutions for marginally challenging environments where auscultation is performed in quiet or controlled rooms with little or no ambient noise.[5,6] Addressing noise contaminations during auscultation is a crucial factor for efficient signal diagnostics. Signal characteristics of natural, environmental noise can yield critical overlap with both normal and abnormal breaths and impair their clinical value:[7] it is only after eliminating interfering noise that disease-diagnostic algorithms can be effectively applied. However, there has been limited relevant work for suppressing real ambient noise in auscultation settings[8,9] . De-noising techniques introduced over the last decade resort on simple filtering methods[10] or use simulated noise environments.[11] Adaptive noise cancellation techniques have been proven inefficient for extreme or unpredicted noise, while the use of alternative, accelerometer sensors has recently been found more promising.[9] Other signal enhancement techniques use the term noise to refer to unwanted signal components, such as suppressing heart beat sounds from a lung sound signal,[12] or separating abnormal explosive occur-

rences from normal airflow.[13] By design, most available studies cannot yet address ambient noise at its fullest or have limited or unknown applicability to generalized clinical settings. Extending results from existing studies to realistic settings is a nontrivial task. If we briefly disregard the importance of noise suppression, available work on disease diagnostics still cannot be generalized in a straightforward manner, since the development of computerized solutions for complete patient diagnostics hasnt yet reached its full potential. Feature extraction and decision-making techniques are developed, applied and validated using supervised, pre-annotated excerpts of data, or manually extracted breaths.[14,15] And, although, this is a very critical part of the processing pipeline before generalizing to global solutions, the ultimate step toward full patient diagnosis has not yet been taken. What would be the optimal way of combining isolated localized results to form a lung disease diagnostic protocol? What is the appropriate way of interpreting segmented lung sound findings still remains unexplored. Furthermore, can such diagnostic protocols be applicable to different auscultation scenarios? How can methods developed with specific assumptions on the auscultation protocol or the target population, be extended to account for the general case. We hope to stimulate interest and invested efforts towards applicable, generalized protocol solutions.

Data availability

Similar to a health care trainee, an automated computer-based algorithm requires large pools of data for training and learning purposes. A greater availability of online breath-sound databases is needed, to incorporate data from everyday settings, in both quiet and busy environments, with no restriction on the characteristics of the subject population. Such diversified data availability will give researchers the opportunity to develop methods applicable to a variety of patients and clinical scenarios, as well as to cross-validate their results, and develop better systems to recognize abnormalities and potential health threats in future occasions. Unfortunately most studies are challenged by limited data availability and an inherent difficulty to evaluate their systems in true

environments. Authors in[16] presented a 97% Positive Predictive Value for their crackle identification model, but the evaluation process was limited to ten lung sound segments of 200 ms. Similar validation methods on only a few breaths were used by most studies.[17–19]

But it is not easy to find publicly available data for the purpose: only a few lung sound databases exist and are available to the scientific community: R.A.L.E,[20] S.T.A.R,[21] and book inserts[1,22] are some of the few examples. Data included within these sets are mostly limited to an adult population auscultated in quiet examination rooms, offering low diversity in lung sounds within normal or abnormal groups. Naturally, automated methods developed and tested after these databases[23] are subject to potential failure when challenged by the busy clinical settings of an ER or by different patient profiles such as pediatric populations. Unlike adults, children cannot be easily instructed on how to behave during a medical examination; typical instructions for slow, steady, deep breathing or restful seating cannot be guaranteed; instead, crying, agitation and fidgeting are a common theme during pediatric examination. Fig. 1.1 illustrates some key differences on the acquired lung sounds: an adult breath sound from a CD-insert database[1] is compared to a pediatric breath recorded in a busy clinic. Arrows indicate abnormal breath patterns while circles depict overlapping energy patterns of ambient noise. Notice the extent of masking introduced by the background crying

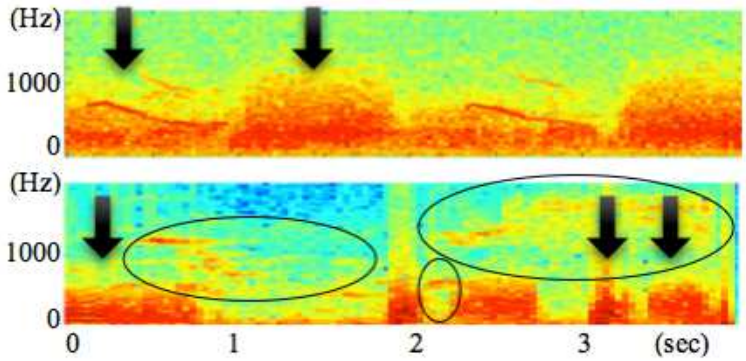

Figure 1.1: Abnormal lung sound spectrograms; top: an adult case from CD-insert database;[1] bottom: a child case recorded in a busy clinic in Zambia. Arrows demonstrate abnormal breaths while circles the ambient noise overlap. Notice the clean, deep, long, steady breaths of the adult case, similar to most lung sound recordings in available databases, and contrast with the highly irregular, explosive breaths of the child case, where background noise is prominently interfering.

and talking in the bottom panel. This is just a glimpse on the inherent difficulty when diagnosing respiratory diseases in a busy, real-life clinical environment. And as mentioned above, it is busy settings like these that have a real need of automated auscultation diagnosis. When auscultating patients, their age group is an important piece of information for physicians; it is, also, crucial to know their medical history, their living conditions and possible geographic risk factors for a more accurate diagnosis. This information is rarely available in volumes, to trainees learning from medical databases. Their rare availability further impacts computer algorithms geared towards automated diagnostic procedures and hampers their adaptability, sensitivity and detection accuracy. In order to expand the impact of the developed methods worldwide, new, extensive, and diversified databases need to become available to medical researchers.

Inherent Uncertainty

Analyzing auscultation signals and developing automated decision-making processes requires intelligent systems that go beyond a yes or no answer, both for easy- and hard-to-diagnose patient cases. Hard-to-make decisions and diagnostic uncertainty cases should ideally be part of a complete, devel-

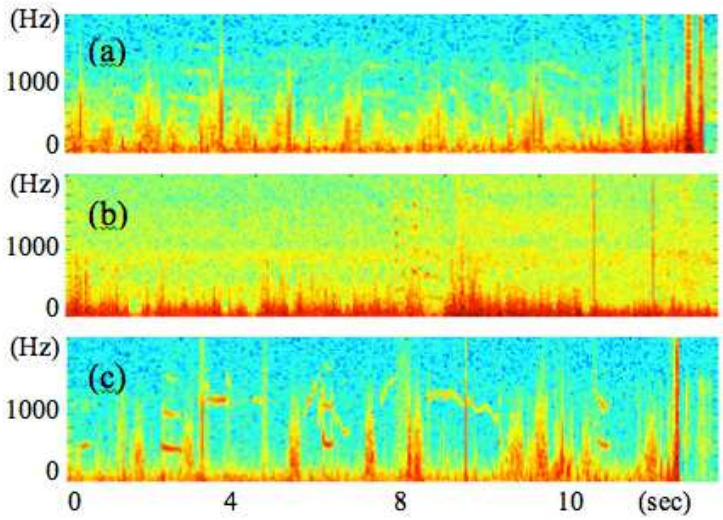

Figure 1.2: Annotated lung sound excerpts from[2] depicting: a) agreement between reviewers; b) uncertain/non-interpretable annotation by at least one reviewer; c) disagreement between reviewers.

oped system. The system can learn how to handle challenging scenarios by incorporating elaborate case examples from real practices. There are currently no standardized databases that can provide the necessary in-depth information. Consider the scenario of an expert panel of two physicians, annotating digital lung sound recordings, possibly for training future practitioners. Fig. 1.2a shows a spectrogram example of a lung sound excerpt where both reviewers agree on the existence of crackles, with certainty, provided by the PERCH study.[2] A different excerpt is shown in Fig. 1.2b, where one of the reviewers indicated that the sound was non-interpretable and that it could not be annotated with certainty, possibly due to the background noise or the quality of the recording. Fig. 1.2c depicts another commonly occurring scenario: disagreement among expert listeners. One reviewer indicated the existence of crackles, while the other indicated a normal breath. When situations like these arise, situations of uncertainty or disagreement, the opinion of a third expert can be requested, or a panel discussion can take place before reaching a final decision. A computer algorithm trained to detect or interpret abnormal sounds can face the same uncertainty. The computerized analysis might reach a grey area where none of the normal or abnormal labels it was trained on can be assigned with confidence. Instead of applying a hard decision, as is typically the case, the system can learn from real practice examples, flag a particular segment as an uncertain case, and request the further attention of a physician or request more incoming data if no assistance is available. Instead of focusing on optimizing hard-decision systems, we need to understand that it is okay for a system to be unsure and we should shift our efforts towards more realistic fuzzy decisions. To our knowledge, there are no studies currently addressing this common matter of uncertain occurrences; and in fact this can hardly be achieved when considering the limited information provided by the available databases. Uncertain or ambiguous interpretation cases are not part of databases, but would be an invaluable added feature when updating or building new ones. Handling hard-to-diagnose cases is an essential part of medical training and, similarly, it should be include in all realistic automated system; the limitations behind a state of the art machine-learning algorithm can very well be due to the limitation of its intrinsic learning procedures and associated learning data.

1.3 Proposed Work and its Significance

We introduce an integrated scheme shown in Fig.1.3 that (i) encompasses noise suppression to improve the signal quality, (ii) offers a rich feature representation to address the unpredictable nature of adventitious auscultation patterns, and (iii) provides patient-level assessment of pathological status by combining partial signal-level assessments without the need for exhaustively detailed annotations. For validation and evaluation, we use a large realistic dataset collected in developing countries in non-ideal rural and outpatient clinics. When it comes to distinguishing between normal vs. pathological lung sounds, we demonstrate the need for noise-free quality signals by using objective quality measures; we further demonstrate the advantages of the proposed feature extraction against state-of-the-art methods, which are shown here to lack the robustness to perform effectively on a diverse set of adventitious sounds, especially when noise events further mask the signal signatures.

Chapter 2 provides details on our methodological approach for improving the quality of the delivered signal, including large scale validation and discussions on state of the art technologies. Chapter 3 presents our approach to providing an improved feature space for auscultation data, able to capture the intricate details of the breath sounds and differentiate normal from abnormal breaths; and Chapter 4 concludes this work, by summarizing the importance of the findings, our continuing efforts, along with directions for future work.

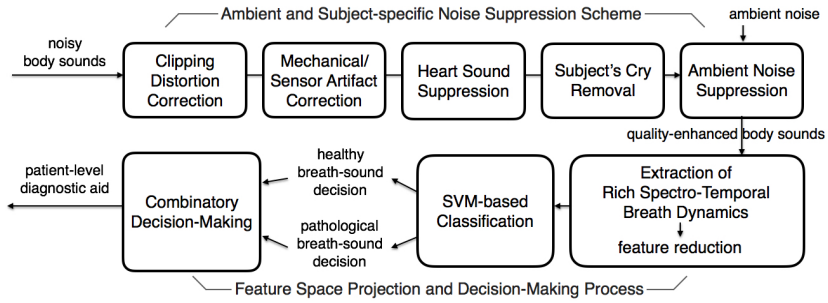

Figure 1.3: Proposed integrated framework for complete auscultation solutions

Significance

Over the last decades, there has been extensive work on computerized analysis of auscultation sounds; but the true potential of automated computer methods has not yet been reached. Due to limited data availability, current studies are developed and evaluated based on limited and isolated breath sound segments, unable to address the true need for full patient diagnostics. And although challenging auscultation protocols or environmental noise are common and can highly impede a physicians work, there is still a large literature gap towards providing robust realistic noise suppression solutions. Bridging the gap between the real need in auscultation care and the available literature would involve an extensive global collaboration between doctors and researchers. Access to diverse database resources would allow incorporation of challenging data and decision-making knowledge into the computer models. This knowledge will enable systems to improve and adapt to unpredicted scenarios. Computerized solutions are far from being a means to substituting medical personnel: once developed to their full potential, such algorithms will be the means to a globally accessible health care. Remote locations and societies depending on limited or minimally trained health care providers will highly benefit from automated diagnostic aid-tools. The required time per patient will be minimized and a more confident diagnosis will be provided. Finally, automated segmentation and classification algorithms can allow enrichment on available data, ensuring a complete, profound training for medical personnel worldwide.

Chapter 2

Obtaining a High Quality Auscultation Signal

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 physicians 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 chapter focuses on 1) characterizing the different types of signal contaminations, expected to be encountered during lung sound measurements in non-ideal environments; 2) a proposed automated multiband denoising scheme for improving the quality of auscultation signals against such heavy background contaminations. *Methods*: Using a database of noise sounds acquired by our collaborating doctors and physicians in a busy hospital in Peru, all encountered noise types were considered, including background talk, radio playing, subjects crying, electronic interference sounds and stethoscope displacement artifacts. The individual characteristics were extracted, discussed and further compared to characteristics of clean segments.

Once the noise profiles were characterized, a noise suppression method was developed to address the heavy contamination of recordings acquired in West Africa and other Asian countries. The proposed noise suppression algorithm works on a simple two-microphone setup, dynamically adapts to the background noise and suppresses contaminations while successfully preserving the lung sound content. The proposed scheme is refined to offset maximal noise suppression against maintaining the integrity of the lung signal, particularly its unknown adventitious components that provide the most informative diagnostic value during lung pathology. *Significance*: Incorporating knowledge of the recommended 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. When it comes to the proposed method for suppressing unwanted contamination, its strengths and benefits lie in the simple automated setup and its adaptive nature, both fundamental conditions for everyday clinical applicability. It can be simply extended to a real-time implementation, and integrated with lung sound acquisition protocols.

2.1 Profiling Noise Contamination in Lung Sound Recordings

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.[24] 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.[25] 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.

2.1.1 Methods & Implementation

Data were obtained from a pool of lung sound recordings acquired in a childrens hospital in Lima, Peru. More information on the acquisition protocols can be found in.[26] 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, $Clean_B$, in-

cluded 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.

Noise profiling

- Spectral Characteristics

  The short-time 214-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

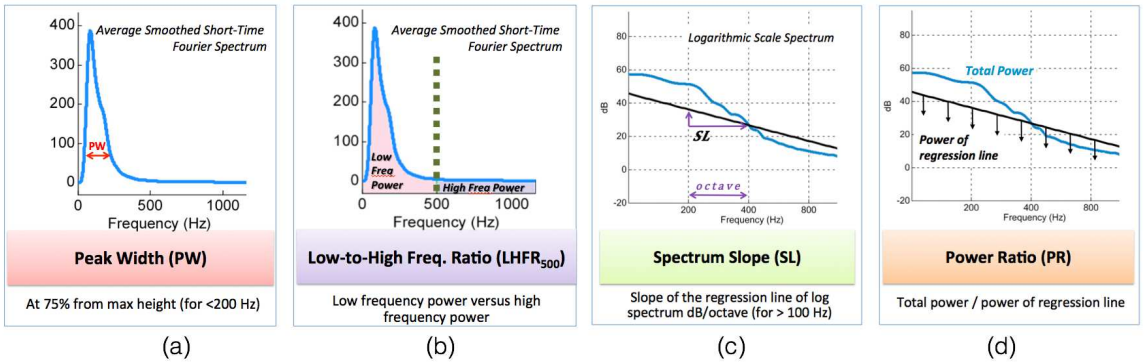

Figure 2.1: Illustration of the spectral characteristics.

(see Fig. 2.1):

- 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. 2.4), the search for the maximum peak was restrained to frequencies below 200 Hz.
- Spectrum Slope (SL) It has been previously shown[27] 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 $20log(P/P_{thr})$, with P_{thr} = 510-5, 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 $P_est(f) = P_{thr}(f/f_{max})^{SL}$, with f_{max} the point where the logarithmic spectrum curve crosses the frequency axis, as proposed in.[27] The power of the regression line depicts the area underneath the linear regression line described above.
- Low-to-High Frequency Ratio ($LHFR_500$) It is 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.[28,29]

- 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 (Fig. 2.2): 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 29-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.

2.1.2 Findings

- Spectral Characteristics

  Twenty reference samples were considered for each one of the five noise sound categories. Segments were processed into short time windows, as discussed earlier, to extract the individual

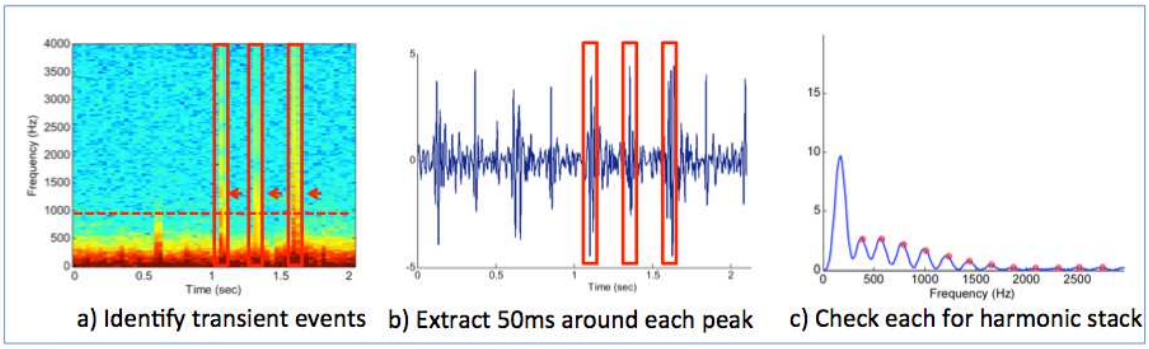

Figure 2.2: Illustration of the harmonicity extraction.

spectrum characteristics. The mean spectrum profile of each class is shown in Fig. 2.4 (left column) and representative examples and spectrum slope plots in Fig. 2.4 (right column). Mean feature values for each group are reported in the Table of Fig. 2.1.2. 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, 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, 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.

Group	Spectrum Features *			
	PW	***SL_{100}***	***PR***	***$LHFR_{500}$***
$Clean_B$	136.65	-9.14	2534.12	27857.58
$Background_N$	162.72	-10.06	10156.26	14702.98
Cry_N	209.28	-10.06	7055.07	2581.76
$Interference_N$	163.70	-9.76	5141.87	9140.24
$StethMove_N$	116.63	-11.78	26453.94	8254.19

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

Figure 2.3: Table showing average spectrum features per sound group.

- Harmonicity

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

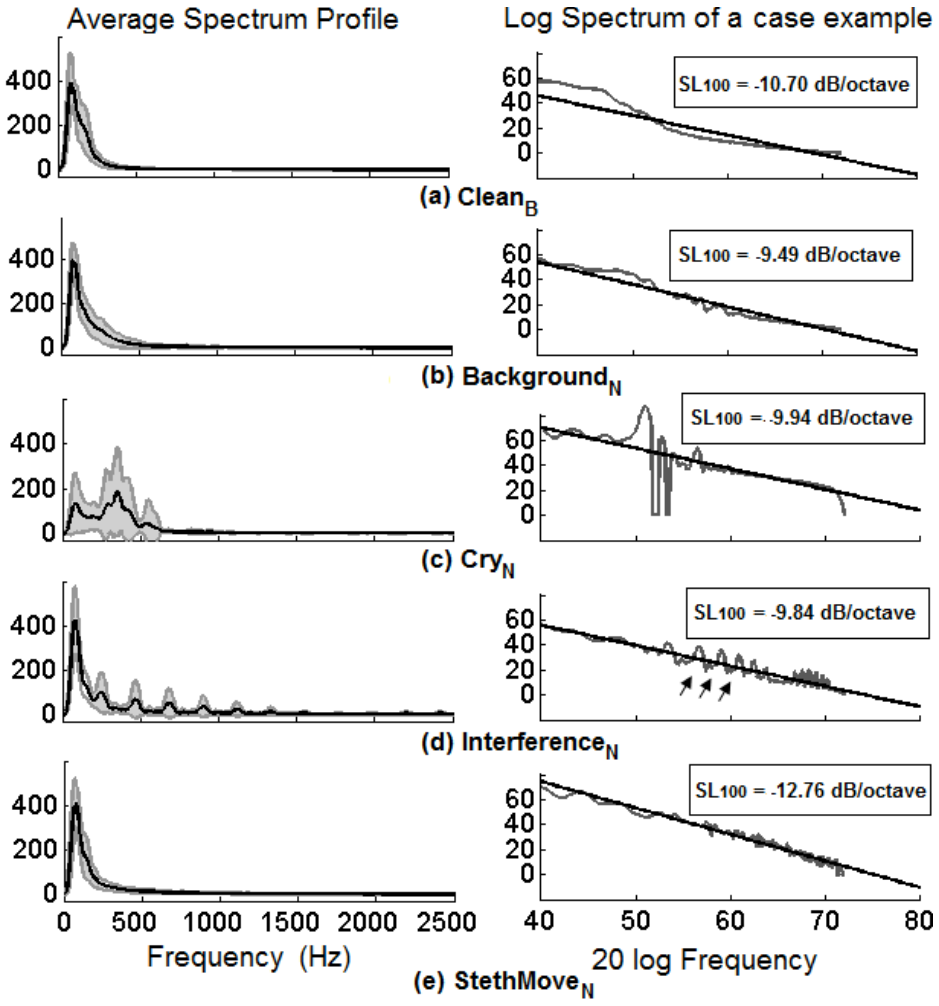

Figure 2.4: Left panels: average spectrum profile of all sound group. Shaded regions reflect the standard deviation among group cases. Right panels: logarithmic spectrum plot of selected case examples. The slope line represents the linear line fit to the spectrum and the slope shown in legend.

these burst of energy in samples of electronic interference a certain musicality emerged, an attribute of signals harmonicity. Such a characteristic is neither heard nor expected for sounds in the $StethMove_N$ group. The reader is referred to Fig. 2.4(d) where arrows indicate the evident harmonic profile of a $StethMove_N$ case. The harmonically detection algorithm 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 $StethMove_N$. Fig. 2.5 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.

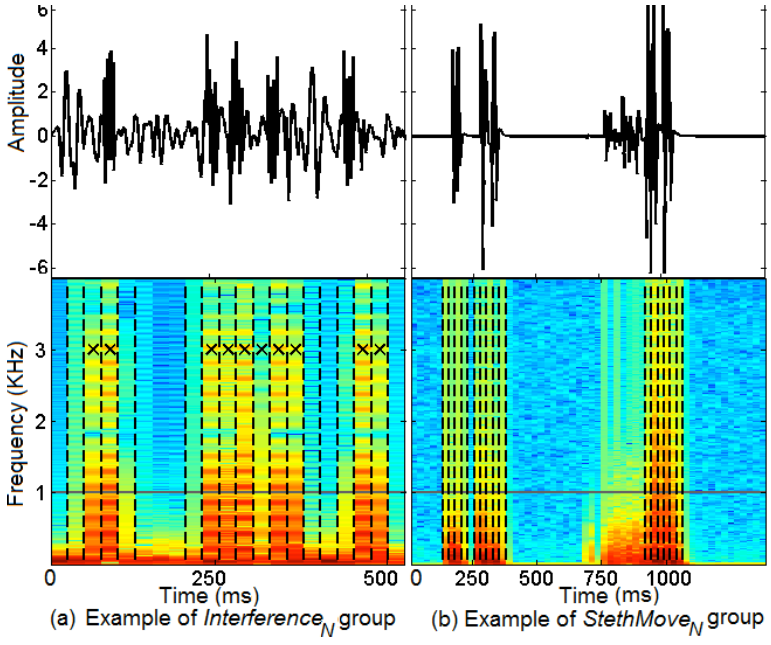

Figure 2.5: Selected case examples of $Interference_N$ (a) and $StethMove_N$ (b) groups. The time waveforms (top panels) and corresponding spectrograms (bottom panels) are shown. Black dashed lines mark the identified transient events of broadband energy. Segments found to exhibit a harmonic structure are noted with an "X" mark.

Discussion

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. 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.

2.2 Improving the Quality of Measured Signal

The issue of environmental and artifact noise contaminations is of particular interest, especially in busy clinics and rural health centers where a quiet examination environment is often not possible, background chatter and other environmental noises are common and patient agitation (especially in children) contaminate the sound signal picked up by the stethoscope. This distortion affects the clarity of the lung sound, hence limiting its clinical value for the health care practitioner. It also impedes the use of electronic auscultation combined with computerized lung sound analysis which are gaining traction in an effort to remedy the inconsistency limitations of standard (acoustic) stethoscope devices and to provide an objective and standardized interpretation of lung sounds.[14,30,31] However, these automated approaches have mainly been validated in well-controlled or quiet clinical settings with adult subjects. The presence of noise impedes the applicability of these algorithms or leads to unwanted false positives.[32]

The current study investigates a series of steps to help alleviate subject-centric and artifact noise. It further proposes the use of multiband spectral subtraction to address noise contaminations in busy patient-care settings where prominent subject-centric noise and room sounds corrupt the recorded signal and mask the lung sound of interest. The setup employs a simple digital stetho-

scope with a mounted external microphone capturing the concurrent environmental or room noise. The algorithm focuses on two parallel tasks: (i) suppress the surrounding noise; (ii) preserve the valuable lung sound content. While spectral subtraction is a generic signal denoising approach, its applicability to the problem at hand is non-trivial in two ways: Firstly, although the signal of interest (i.e. lung sounds) has relatively well defined characteristics,[27,33] unknown anomalous sound patterns reflecting lung pathology complicate the analysis of the obtained signal. These adventitious patterns vary from quasi-stationary events such as wheezes to highly transient sounds such as crackles.[34,35] They are unpredictable irregular patterns whose signal characteristics are not well defined in the literature.[15,36,37] Yet, any processing needs to faithfully preserve these occurrences given their presumed clinical and diagnostic significance. Secondly, noise is highly non-stationary and its signal characteristics differ in the degree of overlap with the signal of interest. Noise contaminations can include environmental sounds picked up in the examination room (chatter, phones ringing, fans, etc.), patient-specific noises (child cry, vocalizations, agitation) or electronic/mechanical noise (stethoscope movement, mobile interference).

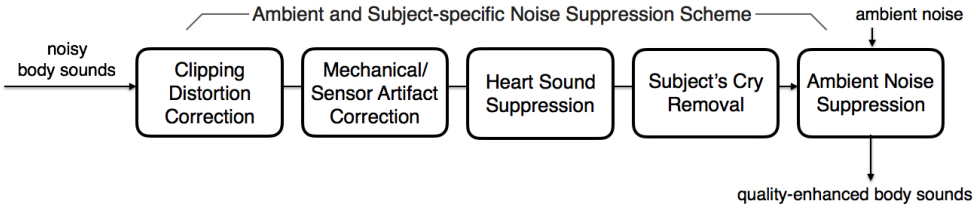

Figure 2.6: Proposed noise suppression scheme for digital auscultation data.

This work tries to balance the suppression of the undesired noise contaminations while maintaining the integrity of the lung signal along with its adventitious components. The proposed scheme is shown in Fig. 2.6. The performance of the proposed approach is validated by formal listening tests performed by a panel of licensed physicians as well as objective metrics assessing the quality of the processed signal.

Clipping distortions

Clipping distortions are produced when the allowed amplitude range of the stethoscope sensor or recording device is exceeded. The incoming sound signal is then truncated, enforcing the loss of high amplitude content and resulting in significant distortion. Both the time and spectral signal signatures are heavily affected by the non-trivial high frequency harmonics formed.

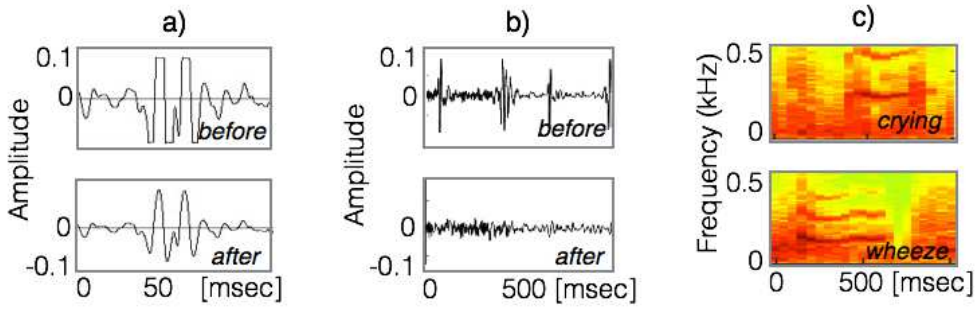

Figure 2.7: (a) Waveform of a lung sound excerpt distorted by clipping (flat amplitude regions in panel "before"), and the corresponding output of the correction algorithm (panel "after"); (b) waveform of a lung sound excerpt illustrating the effects of the heart sound interference suppression; notice the suppressed heart sound patterns (panel "after") when compared to the original waveform ("before"); (c) two spectrogram representations of lung sound excerpts illustrating the inherent difficulty in differentiating between wheezing patterns and crying contamination.

Clipped regions were identified as consecutive time samples with constant maximum-value amplitude, up to a small 3% perturbation tolerance (Fig. 2.7a). Then, the identified regions were repaired using spline piece-wise cubic interpolation; given the brief duration of clipping intervals (a few consecutive data samples), this method was adequate for replacing the distorted portions without distorting the physiological sound signal.

Mechanical or sensor noise is usually generated when the physician moves the stethoscope to various body locations or when the stethoscope is unintentionally and abruptly displaced. This is a common distortion, and especially prominent during pediatric auscultation. Sharp stethoscope movements are typically associated with skin friction and produce irregular short-time broadband energy bursts in the sound signal, resembling profiles of abnormal lung sounds such as crackles. In the current dataset, the stethoscope transition noise was identified as follows: the auditory spectrogram (ASP) representation was calculated on an 8 ms window (described in details later in (3.1)), and normal-

ized to [0,1]. Mostly interested in broadband events, the region of interest ROI_{ASP} within the ASP spectrum, was defined as high spectral content above 1 kHz, with a span greater than 1.5 kHz. Consecutive frames, of 8 up to 100 ms, exhibiting high energy content within ROI_{ASP} were identified and discarded.

In the context of auscultation recordings, heart sounds (HS) are yet another added component masking respiratory sounds. Heart signal suppression has been addressed in several studies using various techniques including wavelets and Short Time Fourier Analysis.[38,39] In order to maintain the integrity of the lung sounds, particularly any adventitious events, a conservative approach was used here, utilizing a wavelet multi-scale decomposition.[40]

(i) HS identification: The original lung sound signal was band-pass filtered in [50, 250] Hz and down-sampled to 1 kHz, using a 4th order Butterworth filter. This step enhanced heart beat components by suppressing lung sounds and noise components outside this range. Next, the discrete Static Wavelet Transform (SWT) was obtained at depth 3, using Symlet decomposition filters (due to their appropriate shape): after Detail $D_j(t)$, and Approximation $A_j(t)$ coefficients were obtained, signals did not undergo down-sampling, which allows for the time-invariance of the transform. Signal reconstruction was then easily obtained by averaging the inverse wavelet transforms.[41] Let $SWT_j\{s(t)\}$ be the wavelet decomposition at the jth scale level of the lung sound signal $s(t)$ and $A_j(t)$ be the obtained normalized approximation coefficient. Then $P_{1:J}(t)$ is the multiscale product of all J approximation coefficients, defined in (2.1). Intervals achieving high values for $P_{i:j}$, were identified as heart sounds and were replaced using an ARMA model.

$$P_{i:j}(t) = \prod_{j=1}^{J} A_j(t)/max(|A_j(t)|) \tag{2.1}$$

(ii) HS replacement: Assuming that lung sounds are locally stationary, an ARMA model was employed to replace missing data of $x(n)$ using past or future values. First a stationarity check

- explained next - was performed on the neighboring area of the removed segment. If the post-neighboring segment was found non stationary, then a forward linear prediction model was used (2a); otherwise, a backward model was used (2b):

$$\hat{x}(n) = -\sum_{k=1}^{p} \alpha_p(k) x(n-k) \tag{2.2a}$$

$$\hat{x}(n-p) = -\sum_{k=0}^{p} \beta_p(k) x(n-k) \tag{2.2b}$$

where $\{-\alpha_p(k), -\beta_p(k)\}$ denote the prediction coefficients of the order-p predictors. Solving for the coefficients by minimizing the mean-square value of the prediction error $\{x(n) - \hat{x}(n)\}$ leads to the normal equations involving the autocorrelation function, $\gamma_{xx}(l)$: $\sum_{k=0}^{p} \alpha_p(k)\gamma_{xx}(l-k) = 0$, with lags $l = 1, 2, .., p$ and coefficient $a_p(0) = 1$. The Levinson-Durbin algorithm was used to efficiently solve the normal equations for the prediction coefficients. The order of each linear prediction model was determined by the length of the particular heart sound gap, using an upper bound of $p_{max} = 125$ ms.

For the stationarity check, the two neighboring intervals around the missing data, of length $T_i = 200$ ms, were partitioned into M non-overlapping windows of length L. Using the Wiener-Khintchine theorem, the power spectral density of the m-th segment, $\Gamma_{xx}^{m}(l)$, was computed via the multitaper periodogram and the following spectral variation measure was introduced[42]

$$V(x) = \frac{1}{ML} \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} (\Gamma_{xx}^{m}(l) - \frac{1}{M} \sum_{k=0}^{M-1} \Gamma_{xx}^{k}(l))^2 \tag{2.3}$$

with $V(x) = 0$ signifying a wide-sense stationary process.

Among identified HS intervals, only the very prominent ones were chosen to be replaced, i.e. the ones achieving increased product values $P_{i:j} > 0.2$. Additionally, if the peak-to-peak interval for identified heart sounds was too short for pediatric standards (< 0.28 s), then the corresponding identified regions (possibly indicative of other adventitious sounds) were not replaced. Fig. 2.7b

shows an example of a heart sound suppressed segment.

Subject's Intense Crying

Depending on the cause of irritation, infants and young children can broadcast crying vocalizations of varying temporal and frequency signature modes:[43,44] phonation, consisting of the common cry with a harmonic structure and a fundamental frequency ranging in 350-750 Hz; hyperphonation, a sign of major distress or pain, also harmonically structured but with rapidly changing resonance and a shifted fundamental frequency of 1-2 kHz or higher; and dysphonation (beyond the scope of this work), a sign of poor control of the respiratory cycle, containing aperiodic vibrations.

Because of their spectral span and harmonic structure, instances of phonation and hyperphonation cry were identified using properties of the signal's time-frequency representation. However, since adventitious lung sounds (particularly wheezes) can produce patterns of similar or overlapping specifications (Fig. 2.7c), here the focus was on longer, intense crying intervals bearing limited value for clinical assessment.

For the detection of phonation mode cry: (i) The ASP representation was calculated for every 8 ms frame (described in details later in (3.1)). A pitch estimate for every frame was calculated, using an adaptation of a template matching approach.[45] Each spectrogram slice was compared to an array of pitch spectral templates, generated by harmonically-related sinusoids, modulated by a Gaussian envelope. The dominant pitch per frame was then extracted and the average pitch (excluding 20% of distribution tails) constituted the resulting pitch estimation per region. Frames with an extracted pitch lower than 250 Hz were immediately rejected. To avoid confusion with possible adventitious occurrences during inspiration or expiration, an identified interval was required to be of duration $T_{dur} > 600\ ms$, considering respiratory rate standards for infants;[46] typical rates in the current dataset were 18 - 60 breaths per minute. (ii) Features of spectro-temporal dynamics (3.1)-(3.5) were extracted from all candidate time-segments, and fed to a pre-trained, binary SVM classifier using radial basis functions, to distinguish crying from other voiced adventitious sounds

like wheezes.

For hyperphonation, simpler steps were required as lung sounds were unlikely to overlap with this type of cry: regions with high ASP spectral content above 1 kHz, and exceeding a duration of T_{dur}, were detected as hyperphonation cry.

In total, 20% of all recorded lung signals were identified as phonation or hyperphonation cry, demonstrating the necessity of such processing step.

Multiband Spectral Subtraction

Spectral subtraction algorithms have been widely used in fields of communication and speech enhancement to suppress noise contaminations in acoustic signals.[47,48] The general framework behind these noise reduction schemes can be summarized as follows: let $y(n)$ be a known measured acoustic signal of length N and assume it comprises of two additive components $x(n)$ and $d(n)$, corresponding respectively to a clean unknown signal we wish to estimate and an inherent noise component which is typically not known. In many speech applications, the noise distortion is estimated from silent periods of the speech signal that are identified using a voice activity detector.[48] Alternatively, the noise distortion can be estimated using a dual or multi-microphone setup where a secondary microphone picks up an approximate estimate of the noise contaminant. Here we employ the latter, a dual-microphone setup capturing both the *internal* signal coming from the stethoscope itself,

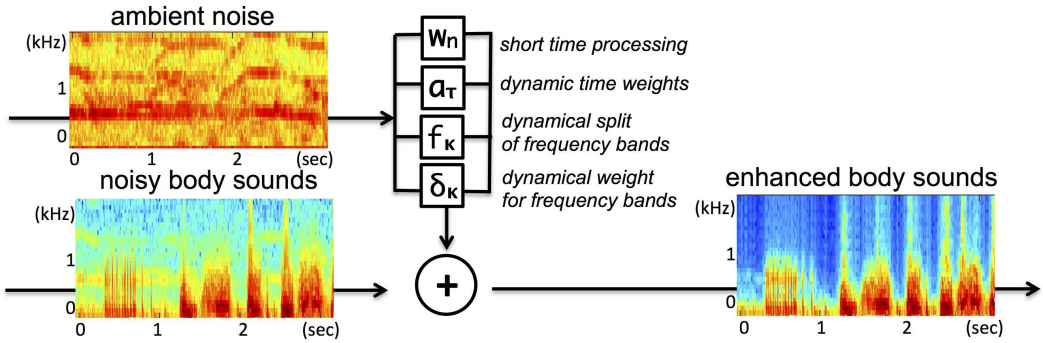

Figure 2.8: Pipeline illustration of the ambient noise suppression scheme.

and the *external* signal coming from a mounted microphone. The external signal is assumed to be closely related to the actual noise that contaminates the lung signal of interest, and shares its spectral magnitude characteristics with possibly different phase profiles due to their divergent traveling trajectories to the pickup microphones. See 2.8 for an illustration.

Noise here is assumed to have additive effects on the desired signal and originate through a wide-sense stationary process. Without loss of continuity, we alleviate the stationarity requirements for the noise process, and assume a smoothly varying process whose spectral characteristics change gradually over successive short-time periods. In this work, such noise signal $d(n,\tau)$ represents the patient or room specific noise signal; $x(n,\tau)$ denotes the desired unknown clean lung sound information, free of noise contaminations; and $y(n,\tau)$ the acoustic information captured by the digital stethoscope:

$$y(n,\tau) = x(n,\tau) + d(n,\tau) \tag{2.4}$$

τ is used to represent processing over short-time windows $w(n)$. In other words, $x(n,\tau) = x(n)w(\tau - n)$ and similarly for $y(n,\tau)$ and $d(n,\tau)$. For the corresponding frequency domain formulation, let $X(\omega,\tau)$ denote the discrete Fourier transform (DFT) of $x(n,\tau)$, implemented by sampling the discrete-time Fourier Transform at uniformly spaced frequencies ω. Letting $Y(\omega,\tau)$ and $D(\omega,\tau)$ be defined in a similar way for $y(n,\tau)$ and $d(n,\tau)$, (2.4) becomes: $|Y(\omega,\tau)|e^{j\phi_y(\omega,\tau)} = |X(\omega,\tau)|e^{j\phi_x(\omega,\tau)} + |D(\omega,\tau)|e^{j\phi_d(\omega,\tau)}$. Short-term magnitude spectrum $|D(\omega,\tau)|$ can be approximated as $|\hat{D}(\omega,\tau)|$ using the signal recorded from the external microphone. Phase spectrum $\phi_d(\omega,\tau)$ can also be reasonably replaced by the phase of the noisy signal $\phi_y(\omega,\tau)$ considering that phase information has minimal effect on signal quality especially at reasonable Signal-to-Noise Ratios (SNR).[49] Therefore, the denoised signal can be formulated as:

$$\hat{X}(\omega,\tau) = \left(|Y(\omega,\tau)| - |\hat{D}(\omega,\tau)|\right) e^{j\phi_y(\omega,\tau)} \tag{2.5}$$

The same formulation can be extended to the power spectral density domain, by making the rea-

sonable assumption that environmental noise $d(n,\tau)$ is a zero-mean process, uncorrelated with the lung signal of interest $x(n,\tau)$:

$$|\hat{X}(\omega,\tau)|^2 = |Y(\omega,\tau)|^2 - |\hat{D}(\omega,\tau)|^2 \tag{2.6}$$

Building on this basic spectral subtraction formulation to synthesize the desired signal, we extend this design in a number of ways:

(i) Extending the subtraction scheme into multiple frequency bands $\{\omega_k\} \in [\omega_k^{min}, \omega_k^{max}]$. This localized frequency treatment is especially crucial given the variable, unpredictable, and non-uniform nature of noise distortions that affect the lung recording (see[7] for a discussion of signal characteristics of noise contaminants). Looking back in equation (2.6), the subtraction term $\hat{D}(\omega,\tau)$ can be weighted differently across frequency bands by constructing appropriate weighting rules (δ_k) that highlight the most informative spectral bands for lung signals.

(ii) Altering the scheme to weight the subtraction operation across time windows and frequency bands, by taking into account the current frame's Signal to Noise Ratio (SNR).

(iii) Reducing the residual noise in the signal reconstruction by smoothing $Y(\omega,\tau)$ estimate over adjacent frames.

Therefore, for frame τ and frequency band ω_k, the enhanced estimated signal spectral density is given by

$$|\hat{X}(\omega_k,\tau)|^2 = |\bar{Y}(\omega_k,\tau)|^2 - \alpha_{k,\tau}\,\delta_k\,|\hat{D}(\omega_k,\tau)|^2 \tag{2.7}$$

Bar notation $\bar{Y}(\omega_k,\tau)$ signifies a smooth estimate of $Y(\omega_k,\tau)$ over adjacent frames. $\alpha_{k,\tau}$ is an over-subtraction factor adjusted by the current frame's SNR, for each band ω_k and frame τ. δ_k is a spectral weighting factor that highlights lower frequencies typically occupied by lung signals[33,50] and penalizes higher frequencies where noise interference can spread. Partial noise is then added back to the signal (2.8) using a weighing factor $\gamma_\tau \in (0,1)$, to suppress musical noise effects.[47,51]

The final estimate $\tilde{x}(n)$ is re-synthesized using the inverse DFT and overlap and add method across frames.[48]

$$|\tilde{X}(\omega_k, \tau)|^2 = (1 - \gamma_\tau)|\hat{X}(\omega_k, \tau)|^2 + \gamma_\tau|\bar{Y}(\omega_k, \tau)|^2 \tag{2.8}$$

2.2.1 Methods & Implementation

Lung signals were acquired using a Thinklabs ds32a digital stethoscope at 44,1 kHz rate, by the Pneumonia Etiology Research for Child Health (PERCH) study group.[2] Thinklabs stethoscopes used for the study were mounted with an independent microphone fixed on the back of the stethoscope head, capturing simultaneous environmental contaminations without any hampering of the physician's examination. Auscultation recordings were obtained from children enrolled into the PERCH study with either World Health Organization-defined severe and very severe clinical pneumonia (cases) or community controls without clinical pneumonia[52] in a busy clinical setting in Basse, Gambia in West Africa. A total of 22 infant recordings among hospitalized pneumonia cases with an average age of 12.2 months (2-37 months) were considered. Following the examination protocol, 9 body locations were auscultated for a duration of 7 s each. The last body location corresponded to a cheek position and is not used in this study.

Noise contaminations were prominent throughout all recordings in the form of ambient noise, mobile buzzing, background chatter, intense subject's crying, musical toys in the waiting room, power-generators, vehicle sirens or animal sounds. Patients were typically seated in their mothers' lap and were quite agitated, adding to the distortion of auscultation signal.

Pre-processing

All acquired signals were low-pass filtered with a 4th order Butterworth filter at 4 kHz cutoff, down-sampled to 8kHz, and centered to zero mean and unit variance. Resampling can be justified by guidelines of the CORSA project of the European Respiratory Society,[50] as lung sounds are mostly concentrated at lower frequencies.

A clipping distortion algorithm was then applied to correct for truncated signal amplitude (occurring when the microphone reached maximum acoustic input). Although clipped regions were of the order of a few samples per instance, they produced very prominent signal distortions. The algorithm identifies regions of constant (clipped) amplitude, and replaces these regions using cubic spline interpolation.[53]

Implementation

The proposed algorithm employs a wide range of parameters that can significantly affect the reconstructed sound quality. An initial evaluation phase using informal testing and visual inspection reduced the parameter space. The preliminary assessment of the algorithm suggests that 32 frequency bands were adequate, using frequency domain windowing to reduce complexity. Since the algorithm operates independently among bands, their boundaries can affect the final sound output. Two ways of creating the subbands were explored: (i) logarithmic spacing along the frequency axis and (ii) equi-energy spacing. The latter spacing corresponds to splitting the frequency axis into band regions containing equal proportions of the total spectral energy. Other band splitting methods were excluded from analysis after the initial assessment phase.

An important factor related to the frequency binning of the spectrum is the weighing among frequency bands, regulated by factor δ_k in (2.7). Since interfering noise affects the spectrum in a non-uniform manner, we imposed this non-linear frequency-dependent subtraction to account for different types of noise. It can be thought of as signal dependent regulator, taking into account the nature of the signal of interest. Lung sounds are complex signals comprised of various components:[50,54,55] normal respiratory sounds typically occupy 50-2500 Hz; tracheal sounds reach energy contents up to 4000 Hz and heart beat sounds vary within 20-150 Hz. Finally, wheeze and crackles, the commonly studied adventitious (abnormal) events, typically have a range of 100-2500 Hz and 100-500 Hz respectively. Other abnormal sounds like stridor, squawk, low-pitched wheeze or cough, all exhibit a frequency profile below 4 kHz. The motivation for appropriately setting factor δ_k is to minimize

distortion of lung sounds that typically occupy low frequencies, and penalize noise occurrences with strong energy content at high frequencies.[7] Our analysis suggested two value sets for parameter δ_k, in Table 2.1. In logarithmic spacing, subbands F_{17}, F_{25}, F_{26}, F_{27} correspond to 80, 650, 850 and 1100 Hz respectively. In equi-energy spacing, F_m corresponds to the m^{th} subband whose frequency ranges are signal dependent; F_{17}, F_{25}, F_{26} roughly correspond to 750, 2000 and 2300 Hz. Comparing the proposed sets, $\delta_k^{(1)}$ resulted in stronger suppression of high frequency content.

Table 2.1: Two proposed sets of values for δ_k

f_k band range	$\delta_k^{(1)}$ value	$\delta_k^{(2)}$ value
$(0, F_{17}]$	0.01	0.01
$(F_{17}, F_{25}]$	0.015	0.02
$(F_{25}, F_{26}]$	0.04	0.05
$(F_{26}, F_{27}]$	0.2	0.7
$else$	0.7	0.7

This non-linear subtraction scheme was further enforced by the frequency dependent over-subtraction factor $\alpha_{k,\tau}$ defined in (2.9) which regulates the amount of subtracted energy for each band, using the current frame's Signal to Noise Ratio. Larger values were subtracted in bands with low a posteriori SNR levels, and the opposite was true for high SNR levels. This way, rapid SNR-level changes among subsequent time frames could be accounted for. On the other hand, such rapid energy changes were not expected to occur within a frequency band, considering the natural environment where recordings took place; thus, the factor $\alpha_{k,\tau}$ could be held constant within bands. Such frame-dependent SNR calculations could also remedy for a type of signal distortion known as musical noise, which can be produced during the enhancement process.

$$\alpha_{k,\tau} = \begin{cases} 4.75 & : \quad SNR_{k,\tau} < -25 \\ 4 - \frac{3SNR_{k,\tau}}{20} & : \ -25 \leq SNR_{k,\tau} \leq \ 40 \\ 1 & : \quad SNR_{k,\tau} > \ 40 \end{cases} \tag{2.9}$$

$$SNR_{k,\tau} = 10\log_{10}(\sum_{\omega\in\omega_k} |\bar{Y}(\omega,\tau)|^2 / \sum_{\omega\in\omega_k} |D(\omega,\tau)|^2)$$

The window length for short-time analysis of the signal was another crucial parameter that can result in noticeable artifacts, since a long time window might violate the stationarity assumptions of the algorithm. Following the initial algorithm assessment phase, we proposed two ways of short-time processing: (i) 50-millisecond window (N=400) and 90% overlap; (ii) 80-milisecond window (N=640) with 80% overlap. Hamming windowing $w(n)$, was applied in the time-waveform to produce all frames. Negative values possibly arising by (2.7) were replaced by a 0.001% fraction of the original noisy signal energy, instead of using hard thresholding techniques like half-wave rectification.

Finally, the enhancement factor γ_τ for frame τ in (2.8) was an SNR-dependent factor and was set closer to 1 for high SNR_τ, and closer to 0 for low SNR_τ values. For the calculation of $\bar{Y}(\omega_k, \tau)$, the smooth magnitude spectrum was obtained by weighting across $\pm$ 2 time frames; given by $|\bar{Y}(\omega_k, \tau)| = \sum_{j=-2}^{2} W(j)|Y_{\tau-j}(\omega_k)|$, with coefficients $W = [0.09, 0.25, 0.32, 0.25, 0.09]$.

Post-processing

Typically, time intervals where the stethoscope is in poor contact with the subject's body tended to exhibit insignificant or highly suppressed spectral energy. After the application of the enhancement algorithm, intervals with negligible energy below 50 Hz were deemed uninformative and removed. A moving average filter smoothed the transition edges.

2.2.2 Validation

The validation of the proposed enhancement algorithm requires a balance of the audio signal quality along with a faithful conservation of the spectral profile of the lung signal. It is also important to consider that clinical diagnosis using stethoscopes is ideally done by a physician or health care professional whose ear has been trained accordingly, i.e. for listening to stethoscope-outputted sounds. Any signal processing to improve quality should not result in undesired signal alterations that stray too far from the 'typical' stethoscope signal, since the human ear will be interpreting the lung sounds at this time. For instance, some aspects of filtering result in "tunnel

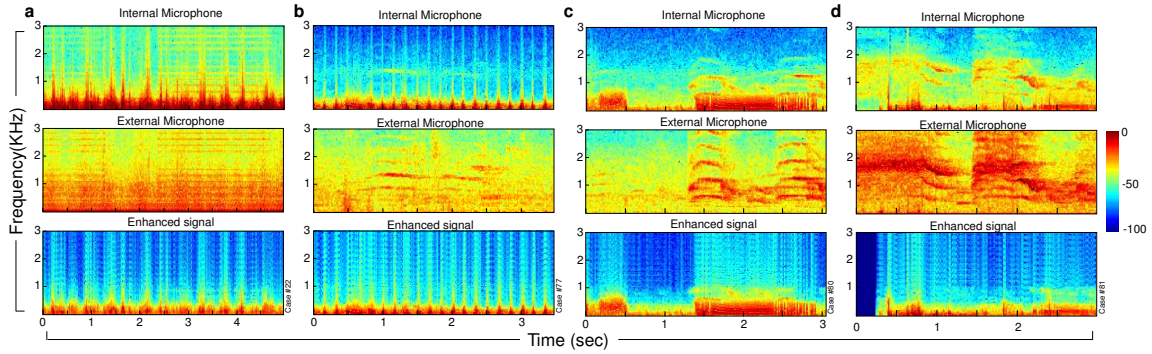

Figure 2.9: Spectrogram representation of four lung sound excerpts. Top panel: internal microphone ; middle panel: external microphone recording; bottom panel: signal as outputted by spectral subtraction algorithm B. The quasi-periodic energy patterns, more pronounced in subfigures (a-b), correspond to the breathing and heart cycles and are well preserved in the enhanced signal. Electronic interference contaminations in (a) and soft background cry in (b) have successfully been removed. Panels (c-d) show cases heavily contaminated by room noise and loud background crying which have substantially been suppressed using the proposed algorithm. Notice how concurring adventitious events were kept intact in (c) at 1.5-3 s and in (d) at 0.6-0.8 s . The period at the beginning of (d) corresponded to an interval of no contact with the child's body and was silenced after the post-processing algorithm.

hearing" effects which would be undesirable even if the quality is maintained. In order to properly assess the performance of the proposed algorithm, we used three forms of evaluations: visual inspection, formal listening tests and objective signal metrics as detailed below. We also used the field recordings employed in the current study to compare performance of existing enhancement algorithms from the literature.

1. Visual inspection

Fig. 2.9 shows the time-frequency profile of four lung sound excerpts appearing per column. Typical energy components that emerge from such spectrograms are the breaths and heart beats, producing repetitive patterns that follow the child's respiratory and heart rate - subfigures (a,b). Such energy components are well-preserved in the enhanced signals (bottom). Middle rows depict concurrent noise distortions captured by the external microphone. Contamination examples include mobile interference (a) and background chatting or crying (b-d) which have successfully been suppressed or eliminated, providing a clearer image of the lung sound energies.

2. Human Listener Experiment

The listening experiment was designed with a two-fold purpose: (i) evaluate the effectiveness of the proposed enhancement procedure and (ii) evaluate the effect of the proposed parameters including frequency band binning, window size and customized band-subtraction factor $\delta_{k,\tau}$ on the perceived sound quality. All methods were designed within the scope of the PERCH study and approved by the Johns Hopkins Bloomberg School of Public Health Institutional Board of Review (IRB).

Participants: Eligible study participants were licensed physicians with significant clinical experience auscultating and interpreting lung sounds from children. A total of 17 physicians (6 pediatric pulmonologists and 11 senior pediatric residents) were enrolled, all affiliated with Johns Hopkins Hospital in Baltimore, MD, with informed consent, as approved by the IRB at the Johns Hopkins Bloomberg School of Public Health, and were compensated for participation.

Data included in the listening experiment was chosen 'pseudo-randomly' from the entire dataset available. Although initial 3 second segments were chosen randomly from the entire data pool, the final dataset was slightly augmented in order to include: (i) abnormal occurrences comprising of wheeze, crackles or other; (ii) healthy breaths; (iii) abnormal and normal breaths in both low- and high-noise environments. A final selection step ensured that recordings from different body locations were among the tested files.

Setup: The experiment took place in a quiet room at Johns Hopkins University and was designed to last for 30 minutes, including rest periods. Data recorded in the field in the Gambia clinic were played back on a computer to participants in the listening experiment. Participants were asked to wear a set of Sennheiser PXC 450 headphones and listen to 43 different lung sound excerpts of 3 s duration each. The excerpts originated from 22 distinct patients diagnosed with World Health Organization-defined severe or very severe pneumonia.[52] For each excerpt, the participant was presented with the original unprocessed recording, along with 4 enhanced versions A, B, C, D.

These enhanced lung sounds were obtained by applying the proposed algorithm with different sets of parameter values, as shown in Table 2.2. In order to increase robustness of result findings, the experiment was divided into two groups consisting of 8 and 9 listeners respectively. Each group was presented with a different set of lung sound excerpts, making sure that at least one excerpt from all 22 distinct patients were contained within each set. In order to minimize selection bias, fatigue and concentration effects, the sound excerpts were presented in randomized order for every participant. The list of presented choices was also randomized so that, on the test screen, choice A would not necessarily correspond to algorithmic version A for different sound excerpts, and similarly for choices B, C, and D.

Table 2.2: Implementation details behind algorithms A, B, C, D running on different short-time analysis windows, frequency band splitting and selection of the band-subtraction factor δ_k.

	A	B	C	D
Window (ms)	50	50	50	80
Band Split	log	equi-linear	log	log
Selection δ_k	$\delta_k^{(1)}$	$\delta_k^{(1)}$	$\delta_k^{(2)}$	$\delta_k^{(1)}$

Listeners were given a detailed instruction sheet and presented with one sound segment at

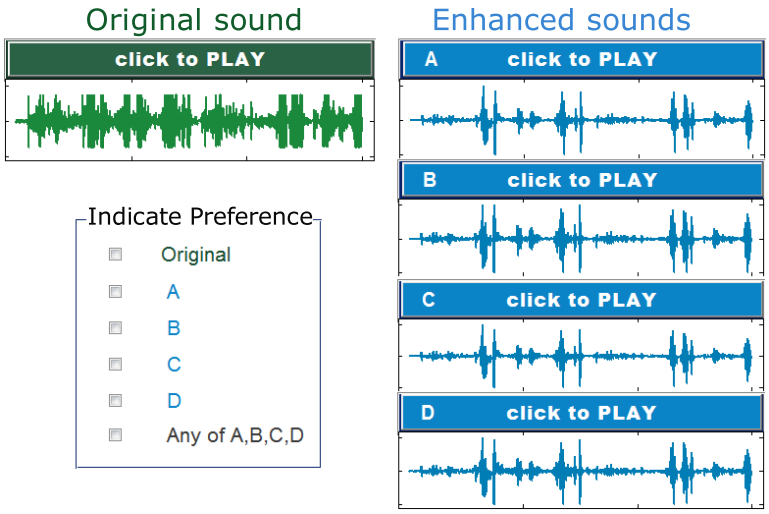

Figure 2.10: Main screen of listening test. Original and enhanced versions are presented for each excerpt before participants indicate their preferred choice.

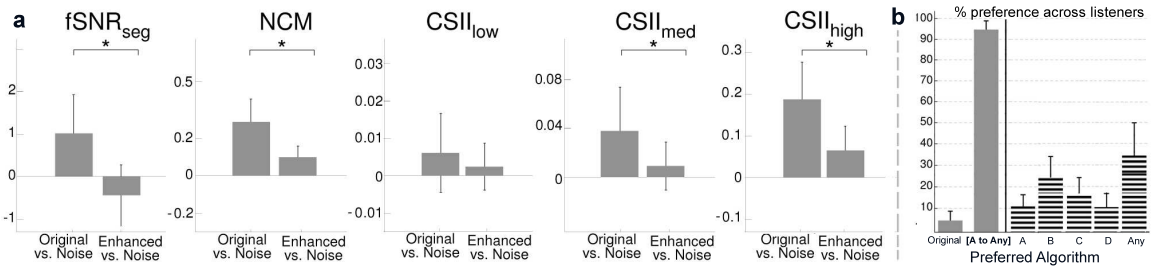

Figure 2.11: (a) Average results with error bars on the evaluation of objective, quality and intelligibility measures for original noisy signal (left bar) and the enhanced signal (right bar), compared with noise as the ground truth. Enhanced signals were found to be more "distant" representations of the noise signals. Stars indicate statistically significant differences. (b) Average responses of the listening text where bars indicate the preference percentage per choice. Left: overall results, comparing average preference of the original sounds versus preference of any of the enhanced versions. Panel *[A to Any]* includes choices {A, B, C, D, Any}; Right: the break-down among all choices. Choice *Any of A,B,C,D* has been abbreviated to *Any*.

a time via a custom designed computer program seen in 2.10. They were asked to listen to each original sound and the enhanced versions as many times as needed. Listeners indicated their preferred choice while considering the preservation or enhancement of lung sound content and breaths, and the perceived sound quality. Instructions clearly stated that this was a subjective listening task with no correct answer. If participants preferred more than one options they were instructed to just choose one of them. If they preferred all of the enhanced versions the same, but better than the original, an extra choice, "Any", (brief for "Any of A,B,C,D") was added.

Results: Fig. 2.11b summarizes the opinions of the panel of experts. Considering all listeners and all tested sound excerpts, the bars indicate the percentage of preference among the available choices. Bar plots were produced by first forming a contingency table per listener, counting his/her choice preferences, and then averaging across listeners. The vertical lines depict the standard variation among all listeners.

The listed choices on the x-axis correspond one by one to the ones presented during the listening test, where choice *Any of A,B,C,D* has been abbreviated to *Any*. An extra panel, *[A to Any]*, is added here illustrating preference percentages for any enhancement version of the algorithm, irrespective of choice of parameters. On average, listeners prefer mostly choice *Any* (34.06 % of the

time), followed by choices B and C. Overall, listeners prefer the enhanced signal relative to the original -unprocessed- signal 95.08 % of the time. Considering responses across groups of listeners, results are consistent across Group 1 and Group 2. A statistical analysis across the two Groups using a parametric t-test and a non-parametric Wilcoxon rank sum test show no difference among the two populations except possibly for choice D. The corresponding p-values for the t-test and the Wilcoxon test, (p_t, p_w), are ; for choice *Original*: (0.28, 0.23); choice A: (0.37, 0.52); choice B: (0.74, 0.62); choice C: (0.33, 0.74); choice D: (0.08, 0.10); choice *Any*: (0.11, 0.05); choice *[A to Any]*: (0.28, 0.23).

Discussion: Analyzing the results, choice C is preferred over B when the test sound consists of a low or fade normal breath. To better understand this preference, it is important to note that algorithm C is relaxed for higher frequencies due to the δ_k parameter. Qualitatively, low-breath excerpts all retained the normal breath information after noise suppression, but with an added soft wind sound effect. This wind distortion or hissing was at a lower frequency range for algorithm B and proved to be less pleasant than the one produced by algorithm C, which ranged in higher frequencies. This observation was consistent across different files and listeners. Looking further into algorithm C, a larger preference variation was noticed for group 2 when compared to group 1. This variation was found to be produced by two participants who preferred C over any other choice 35% of the time and who both preferred the original only in two cases.

The original recording was preferred 4.9% of the time. While this percentage constitutes a minority on the tested cases, a detailed breakdown provides valuable insights on the operation of the enhancement algorithm. In most cases, it is determined that low-volume resulting periods affect the listeners' judgments.

- Clipping distortions make abnormal sound events even more prominent. Clipping tends to corrupt the signal content and produce false abnormal sounds for loud breaths. However, when such clipping occurs during crackle events, it results in more distinct abnormal sounds, which can be better perceived than a processed signal with muted clipping. For two such

sound files in group 1, 2/8 users prefer the original raw audio and for one such file in group 2, 2/9 prefer the original.

- Child vocalization are typically removed after enhancement. Since the algorithm operates with the internal recording as a metric, any sound captured weakly by the internal but strongly by the external microphone is flagged as noise. One such file in group 2 leads 4/9 users to prefer the original sound: a faint child vocalization is highly suppressed in the enhanced signal. As users are not presented with the external recording information, it can be hard to tell the origin of some abnormal sounds that overlap with profiles of abnormal breaths. Nevertheless, a post analysis on the external microphone shows that this is indeed a clear child vocalization.

- Reduced normal breath sounds. The proposed algorithm has an explicit subtractive nature; the recovered signal is thus expected to have lower average energy compared to the original internal recording. Before the listening test all recordings are amplified to the same level; however isolated time periods of the enhanced signal are still expected to have lower amplitude values than the corresponding original segment, especially for noisy backgrounds. This normalization imbalance has perceivable effects in some test files. For auscultation recordings in lower site positions, breath sounds can be faintly heard, and the subtraction process reduces those sounds even further. Two such cases were included in the listening test, where suppression of a loud power generator noise resulted in a faded post-processed breath sound. In this case, listeners preferred the original file where the breath sounds stronger than the processed version.

A finalized enhancement algorithm is proposed consisting of parametric choices that combine versions B and C. The smoother subtraction scheme enforced by factor $\delta_k^{(2)}$ is kept along with the equi-linear model of frequency band-splitting using a 50 ms frame size window. An informal validation by a few members of the original expert panel confirms that the combined algorithm parameters result in improved lung sound quality and preservation of low breaths.

3. Objective validation of processed signals

To further assess improvements on the processed signals, objective methods were used to compare the signals before and after processing. Choosing an evaluation metric for enhancement is a non-trivial issue; many performance- or quality measures commonly proposed in the literature often require knowledge of the true clean signal or some estimate of its statistics.[56] This is not feasible in our current application: bio-signals such as lung sounds have both general characteristics that can be estimated over a population, but also carry individual traits of each patient that should be carefully estimated. It is also important to maintain the adventitious events in the lung sound, while mitigating noise contamination and other distortions. To provide an objective assessment of the proposed method, we employed a number of qualitative and quantitative measures, coming from telecommunication and speech processing fields but adapted to the problem at hand. The metrics were chosen to assess how much shared information remains in the original and enhanced signals, relative to the background noise recording. While it is important to stress that these are not proper measures of signal quality improvement, they provide an informative assessment of shared signal characteristics before and after processing:

* *Segmental Signal-to-Noise Ratio (fSNRseg):* Objective quality measure, estimated over short-time windows accounting for signal dynamics and non stationarity of noise.[48]

$$fSNRseg = \frac{10}{T} \sum_{\tau=1}^{T} \frac{\sum_{k=1}^{K} w_k SNR^F}{\sum_{k=1}^{K} w_k}$$

with $SNR^F = log_{10}\{(|X(k,\tau)|^2) \ / \ (|X(k,\tau)| - |\hat{X}(k,\tau)|)^2\}$, where w_k represents the weight for frequency band k, $\hat{X}$ represents the processed signal, and X typically represents the clean (desired) signal. As mentioned above, in this work X will represent the background noise, since the clean uncontaminated signal in not available. SNR^F is calculated over short-time windows of 30 ms to account for signal dynamics and non stationarity of noise, using a Hanning window. For each frame, the spectral representations $X(k,\tau)$ and $\hat{X}(k,\tau)$ are computed by critical band filtering. The bandwidth and center frequencies of the 25 filters used and the perceptual (Articulation Index)

weights w_k follow the ones proposed in.[48,57] Using the described method, fSNRseg value can reach a maximum of 35 when the signals under comparison are identical. Comparatively, a minimum value just below -8 can be achieved when one of the signals comes from a white Gaussian process.

** Normalized-Covariance measure (NCM):* A metric used specifically for estimated speech intelligibility (SI) by accounting for audibility of the signal at various frequency bands. It is a speech-based Speech Transmission Index (STI) measure capturing a weighted average of a Signal to Noise quantity SNR^N, where the latter is calculated from the covariance of the envelopes of the two signals over different frequency bands k[58] and normalized to [0,1]. The band-importance weights w_k followed ANSI-1997 standards.[59] Though this metric is speech-centric (as many quality measures in the literature), it is constructed to account for audibility characteristics of the human ear hence reflecting a general account of improved quality of a signal as perceived by a human listener.

$$NCM = \{\sum_{k=1}^{K} w_k \; SNR^N(k)\} \; / \; \sum_{k=1}^{K} w_k$$

** Three-level Coherence Speech Intelligibility Index (CSII):* The CSII metric is also a speech intelligibility-based metric, based on the ANSI standard for the Speech Intelligibility Index (SII). Unlike NCM, CSII uses an estimate of Signal-to-Noise ratio in the spectral domain, for each frame $\tau = 1, ..., T$: the signal-to-residual SNR^N_{ESI}; the latter is calculated using the ro-ex filters and the Magnitude-Squared Coherence (MSC) followed by [0,1] normalization. A 30 ms Hanning window was used and the three-level CSII approach divided the signal into low, mid, and high-amplitude regions, using each frame's root mean square level information.[48,60]

$$CSII_{.} = \frac{1}{T} \sum_{\tau=1}^{T} \frac{\sum_{k=1}^{K} w_k SNR^N_{ESI}(k,\tau)}{\sum_{k=1}^{K} w_k}$$

All metrics generally require knowledge of the ground-truth undistorted lung signal, which is not available in our setup. In the current work, we apply them to contrast how much information is shared between the improved and the background (noise) signal, relative to the non-processed (original) auscultation signal. Specifically, each metric was computed between the time-waveforms of the original $y(n)$ and the background noise $\hat{d}(n)$ signals; then contrasted for the enhanced $\tilde{x}(n)$

and the background $\hat{d}(n)$ signals. The higher the achieved metric value, the "closer" the compared signals are, with respect to their sound contents. Fig. 2.11a shows histogram distribution results for each metric: Segmental Signal-to-Noise Ratio (fSNRseg) yielded on average a value of 1.02 between the original and the noise signals, likely reflecting leak through the surrounding environment to the internal microphone. Such measure was reduced to -0.44 when contrasting the improved with the noise signal indicating reduced joint information. The two distributions were statistically significantly different (paired t-test: t-statistic=15.99 and p-value p_t=3E-13; Wilcoxon: Z-statistic=4.5 and p-value p_w=8E-06) providing evidence that the original signal was "closer" -statistically- to the surrounding noise, relative to the enhanced signal. Significant difference was also observed in all other metrics (Fig.2.11a), with NCM (p_t=1E-10; p_w=2E-06) , CSII$_{med}$ (p_t=1E-10; p_w=3E-05) and CSII$_{high}$ (p_t=7E-10; p_w=7E-06).

2.3 Comparison with state-of-the-art methods and technology

This comparion entails two parts: A) comparison of proposed methods with published work on noise suppression for auscultation signals; and B) comparison of novel stethoscope design embedding the proposed methods, with commercially available products.

2.3.1 Part A: Comparison with published work

A proper comparison to existing noise suppression methods for auscultation signals is largely limited due to the scarce literature on this topic, especially when dealing with busy real-life environments, particularly in pediatric patients. Published methods typically consider auscultations in soundproof chambers, highly controlled environments with low ambient or Gaussian noise[5]-.[6] Moreover, the term noise often refers to suppressing heart sounds in the context of healthy lung

sound analysis[61]-,[12] or to separate normal airflow from abnormal explosive occurrences[13]-.[24] Extending results from published studies to realistic settings is nontrivial, particularly in non-healthy patients where abnormal lung events occur in an unpredictable manner and whose signal characteristics may overlap with those of environmental noise.

Here, we contrast our results with the performance of a published lung sound enhancement scheme,[62] which mainly focuses on post-classification of auscultation sounds, rather than production of improved-quality auscultation signals to be used by health care professionals in lieu of the original recording. The authors adopted the speech-based spectral subtractive scheme of Boll, 1979[63] which has well documented shortcomings.[64,65] For a fairer comparison, we used a more robust instantiation of speech-based spectral subtraction, proposed in,[48,66] which we call here *speechSP*. We contrasted our proposed method with *speechSP*, maintaining the same window size, window overlap factor and number of frequency bands as mentioned above; both algorithms were applied on the same pre-processed signals, after downsampling, normalizing and correcting for clipping distortions.

A visual inspection of the *speechSP* method is sufficient to observe the notable resulting artifacts. Fig.2.12 (a) illustrates an example comparing the two methods when applied on the same auscul-

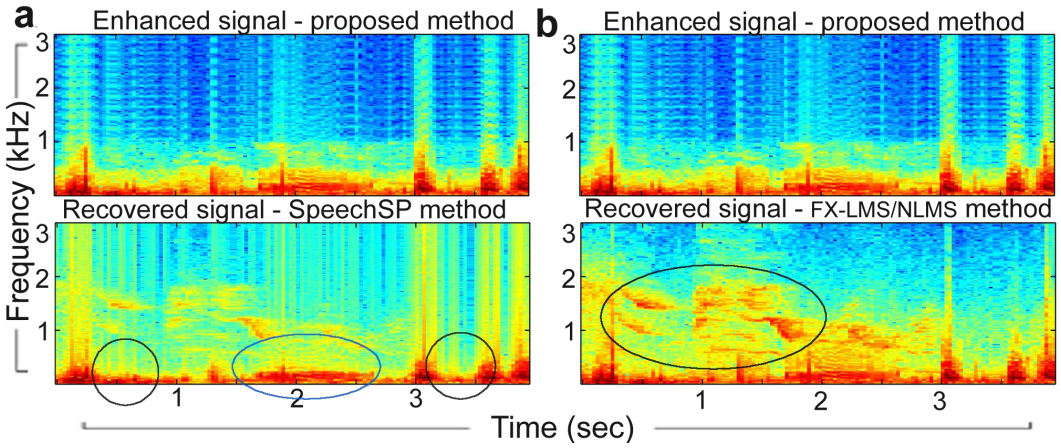

Figure 2.12: Spectrogram illustrations comparing the proposed method with *speechSP* (a), and FX-LMS (b) applied on the same sound excerpt. *SpeechSP* suppresses important lung sounds like crackle patterns (black circles) and wheeze pattern (blue circle). FX-LMS convergence is challenged by both the parametric setup and the complex, abrupt noise environment resulting in non-optimal lung sound recovery. Colormap is the same as Fig.1.

tation excerpt. *SpeechSP* algorithm highly suppressed the wheezing segment around 2 s in Fig. 2.12a, along with the crackle occurrences around 0.5 and 3.5 secs. In this example (and all cases in the current study -not shown-), the *speechSP* method suffered from significant sound deterioration; and in the majority of cases, the *speechSP*-processed signal was corrupted by artifacts impeding the acoustic recognition of alarming adventitious events. Overall, the combination of visual inspection, signal analysis and informal listening tests, clearly indicate that *speechSP* maximizes subtraction of background noise interference, at the expense of deterioration of the original lung signal as well as significant masking of adventitious lung events. Both effects are largely caused by its speech-centric view which considers specific statistical and signal characteristics for the fidelity of speech that do not match the nature of lung signals.

Next, we compared the proposed method to Active Noise Cancellation (ANC) schemes. Such algorithms typically focus on noise reduction using knowledge of a primary signal and at least one reference signal. Here, we consider the case of a single reference sensor and use a feed-forward Filtered-X Least Mean Square algorithm (FX-LMS). FX-LMS has been previously used for denoising in auscultation signals recorded in a controlled acoustic chamber with simulated high noise interference.[67] Here, we adopt an implementation of the Normalized LMS (NLMS) as in[67]-.[9] Using all signals of the current study, we tested the effectiveness of NLMS in suppressing external noise interference. The filter coefficients were optimized in the MSE sense, with filter tap-order N_{LMS} varying between $[4, ..., 120]$, step size η_{LMS} varying between $[1E\text{-}08, ..., 2]$ and denominator term offset step size C_{LMS} in $[1E\text{-}08, ..., 1E\text{-}02]$. A representative example is shown in Fig. 2.12b; zero initial filter weights were assumed with the optimal solution occurring for $(N_{LMS}, \eta_{LMS}, C_{LMS})$=$(90, 5E\text{-}7, 1E\text{-}8)$. Our results indicate that NLMS fails to sufficiently reduce the effect of external noise, especially in low SNR instances or during abrupt transitions in background interferences.

As previously noted in,[9] difficult acoustic environments typically pose a challenge to ANC methods for auscultation where ambient recordings are rendered ineffective as reference signals. This limita-

tion is due to a number of reasons.[68] Firstly, the presence of uncorrelated noise between the primary and reference channels largely affects the convergence of NLMS and the performance of the denoising filter. Nelson et al.[9] have indeed demonstrated that, using an external microphone is sub-optimal in case of auscultation recordings, proposing use of accelerometer-based reference mounted on the stethoscope in line with the transducer; a non-feasible setup for our study. Furthermore, iterative filter updates in the NLMS are heavily dependent on the statistics of the observed signal and reference noise.[69] Abrupt changes in signal statistics pose real challenges in updating filter parameters fast enough to prevent divergence[70].[71] This is particularly true in field auscultation recordings where brusque changes in the signal often occur due to poor body seal of the stethoscope - caused by child movement or change of auscultation site. Noise sources are also abruptly appearing and disappearing from the environment (e.g. sudden patient cry, phone ring); hence posing additional challenges to the convergence of the algorithm without any prior constraints or knowledge about signal statistics or anticipated dynamics. Furthermore, unfavorable initial conditions of the algorithm can highly affect the recovered signal and lead to intractable solutions.

2.3.2 Part B: Comparison with commercially available technology

Here we attempt a comparison with state of the art electronic stethoscopes, while considering a variety of auscultation scenarios. We invented a new screening tool, the X2V, that is able to embed the proposed noise suppression system and put it in action. This novel smart device mitigates a number of limitations in the existing auscultation systems discussed above, while also providing the flexibility to realize solutions to automated body sound recognition as a computerized-aid via the use of add-on embedded software. It allows for real-time capabilities that include active filtering of ambient noise: knowledge of the lung sounds profile and the concurrent noise profile is used to dynamically adapt to abrupt and highly unpredicted environmental noise. It offers the ability to

maintain the sounds of interest in any challenging environ- ment and therefore provides quality care and offers increased confidence in following diagnostic procedures.

The aim here is to evaluate the quality of the body signal that reaches the ears of the end-user of a particular auscultation system. In order to test the capabilities of the proposed auscultation system, a simulated noisy clinical setting was created inside a sound booth. Body sounds were played via a chest sound simulator device (details to follow) placed on top of an examination table. The table was surrounded by loudspeakers that played various noise combinations reflecting a variety of examination settings from a quiet room to a busy clinic. The device was placed ontop of the chest sound simulator, and the auscultated signal was recorded for quality evaluation. The same setup was used to further compare the proposed system with six commercially available auscultation systems incuding both acoustic and electronic devices.

Proposed system

The sensors: The body sensing unit is an array of five omnidirectional microphones from Shenzhen Horn Electroacoustic Technology. The microphones are spaced 7mm apart in a cross pattern with the fifth microphone in the middle. The choice of the array ensures uniform pickup throughout the surface of the chest-piece. An externally-facing omnidirectionl microphone is secured next to the sensing array for capturing environmental sound signals.

The DSP hardware: The digital hardware design comprises of two main sys- tems an audio codec and a microprocessor. The implemented audio codec is the NXP SGTL5000, a low power, 96 kHz, 24-Bit audio codec. The microprocessor is the low-power NXP Kinetis MK64 microcontroller unit with a USB inter- face. These two systems work in tandem to receive the signal from the sensing units and implement a real-time noise- cancellation. The output signal is fed to a standard 3.5mm headphone jack for real-time listening and monitoring. The microprocessor can also store audio data on a micro SD card for data logging in the field or clinic without the need of a secondary device. If needed, it can be re-programmed with updated software and algorithms through the USB port.

The software: The embedded noise-cancellation feature works in real-time and can dynamically adapt to the current temporal and spectral signatures of the noise. Implementation is based on the spectral subtractino methods presented above.

Methods

Clinical Room Simulation: Inside a soundbooth, of dimensions (148in long) x (123in wide) x (89in high), six loud speakers, one examination table and one chest sound simulator were placed, Fig. 2.13. The examination table was stationed towards one end of the room while the six Genelec 6010A loudspeakers were arranged to losely face towards it at different angle and height placements (Table 2.3), independently broadcasting noise sounds of various types and levels. On top of the examination table, the chest sound simulator transmitted low volume body signals. The noise and body sounds were delivered via a connected computer, stationed outside of the room.

Chest Sound Simulation: A chest sound simulator (ChestSim) was designed for transmitting the relevant body sounds, built comprising of a loudspeaker covered in ballistic gelatin. The ChestSim transmitted digital breath signals at a low, fixed level, comparable to real chest auscultation signals

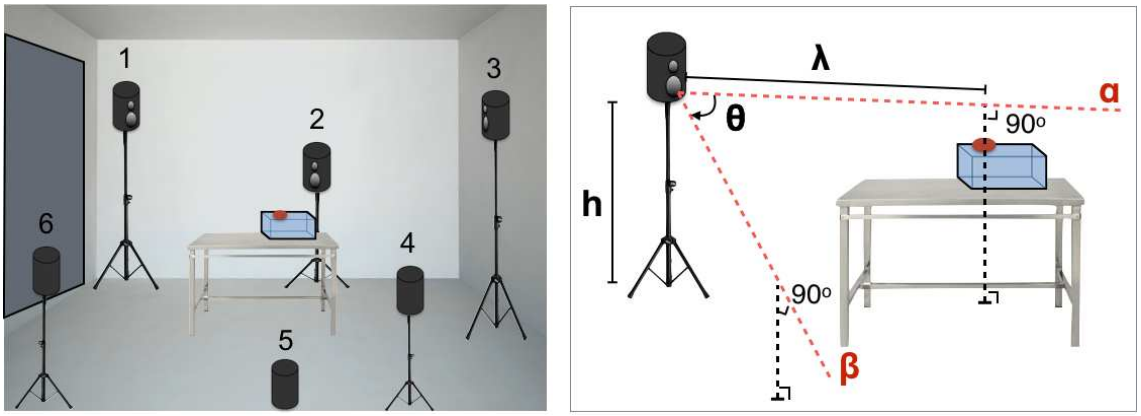

Figure 2.13: Left panel: schematic of the experimental setup illustrating the placement of the loudspeakers and the chest sound simulator (blue rectangular prism). The red circle on the chest simulator illustrates the designated signal pickup area (SPA), used as a reference point for measuring the loudspeakers' relative position. Right panel: illustration of the loudspeaker placement calculation, with individual speaker angle and position shown in Table 2.3.

Table 2.3: Speaker placement relative to the SPA reference point on the chest sound simulator

Speaker:	**1**	**2**	**3**	**4**	**5**	**6**
h (m)	1.55	0.88	1.50	0.76	0.00	0.77
λ (m)	1.51	0.84	1.65	1.75	1.60	0.76
θ (angle ^{o})	+47	0	−9	+30	+6	−9

via a connected computer installed outside of the soundbooth. It was built comprising of a Jawbone Jambox loudspeaker with frequency response of 40-20,000 Hz and improved low-frequency sensitivity (via a proprietary bass radiator). The loudspeaker was covered in 1.5in-thick ballistic gelatin from Clear Ballistics, that closely simulates the density and viscosity of human muscle tissue and can be kept at room temperature without deforming.[72, 73]

Auscultation Process Simulation: The device of an individual auscultation system was placed at the designated SPA area on top of the chest simulator (Fig. 2.13) and held in position using a clamp. The clamp, braced on a pole mount, was secured around the lowest surface of the device, while a moderate pressure was applied as a result of the clamp's weight. This moderate pressure level was preferred over heavier pressure so as to create a more realitic auscultation setup. The setup remained the same throughout the completion of all simulations.

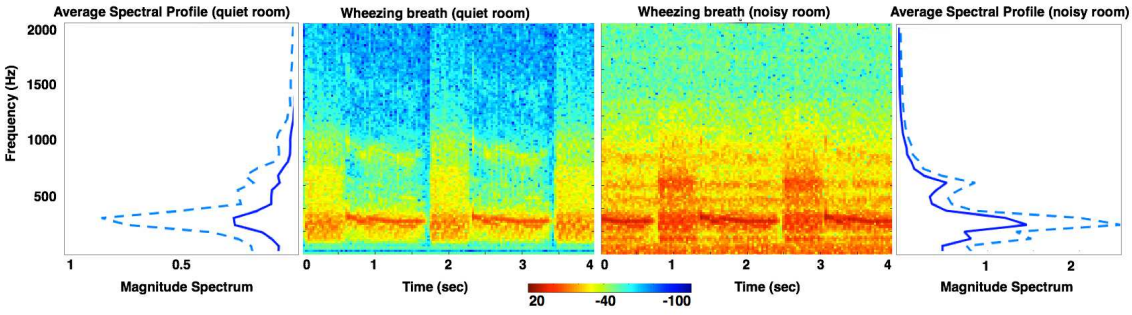

Figure 2.14: Spectrogram plots of a wheezing breath sound in the Abnormal group during quiet auscultation (middle left) and noisy auscultation of -10dB SNR high stationary noise (middle right). The average spectral profile of all lung sounds is shown in the leftmost panel for the quiet condition, and the rightmost panel for the noisy conditions (all net noise recordings at -10dB SNR) respectively. The dashed lines correspond to the upper bound of the standard deviation over all sounds. Mind the different magnitude axis.

Table 2.4: Ambient Noise Database Grouping

High Stationary	Comprised exclusively of various colored noises and fan noise
Low Stationary	Comprised of ambulance sirens, pulse monitor sounds, babble and crying noises, chirp sounds and street noise

Data preparation - Chest sound data: A collection of ten abnormal and ten control breath sounds of 10 s duration each were selected from.[74] The abnormal group consisted of breath sounds containing wheeze, crackle and stridor sounds, while the control group consisted of mostly normal breath sounds recorded over various chest and tracheal areas. All digital clips were downsampled to 8 kHz. Examples taken from the database are shown in Fig. 2.14, along with average spectral profiles per group.

Data preparation - Noise data: The ambient noise database was selected having a variety of clinical settings in mind varying from a quiet clinic to a busy examination room, and sounds were split into two main categories: High Stationary and Low Stationary (see Table 2.4). The **High Stationary** group consisted of several colored noise subgroups including white, pink, violet, blue and brown, and fan-like noise found in the BBC database.[75] In total, 20 ten-second clips were selected from each subgroup, resulting in 120 High Stationary noise sounds. The **Low Stationary** noise group consisted of noise types found in the BBC and NoiseX-92 databases,[75,76] and included subgroups of hospital ICU noise, hospital corridor noise, pulse monitor sounds, ambulance noise, babble noise and ambient talk, baby cry, street noise, chirping birds. Random silence periods were intejected to the noise clips to accentuate their non-stationary nature.

Sound playback and capture: While the ChestSim emitted the selected body sounds at a fixed low volume, the loudspeakers independently broadcasted noise sounds randomely selected from the database. The volume of individual speakers was set to a random difference of $\{0, \pm 0.5, \pm 1, \pm 2\}$ dB from each other. The master speaker volume was automatically adjusted at th ebeginning of each trial, ensuring a net noise effect of Signal-To-Noise ratio (SNR_{true}) varying within $[-20, ..., 15]$ dB. All transmitted lung sound clips were pre-amplified independently to ensure equal

average sound levels. In total 10 normal and 10 control lung sounds were used, each assigned to 5 random net noise combinations (trials), resulting in a total of 100 sound recordings per SNR_{true} value: 50 abnormal and 50 control combination sets. The true SNR level, SNR_{true}, for each breath sound/net noise combination was determined by the short-term average of the ratio of the individual signal powers, averaged over M frames of duration 100 ms each:

$$\widehat{SNR}_{true} = \frac{1}{M} \sum_{m=1}^{M} 10 \ \log_{10} \ \frac{P_s(n)}{P_d(n)} \tag{2.10}$$

where $P_*(n)$ is the average signal power of the n^{th} time frame; s corresponds to the sound signal picked up on the designated signal pickup area (SPA) on top of the chest simulator (Fig 2.13 left); and d corresponds to the net ambient noise picked up adjacent to the designated SPA point. For the calculation of SNR_{true}, only the top 30% frames were considered, achieving the average highest signal power. This way, it is ensured that no sound events exceed the desired SNR, while allowing for lower sound level events to be present.

For the calculation of the true SNR, the signals of interest s and d were obtained using two PCB Piezotronics prepolarized condenser microphones of 1/4" pressure, connected to a Brüel & Kjaer type 5935-L preamplifier at 20dB gain. The first microphone was placed ontop of the ChestSim, facing downwards onto point SPA, for recording signal s; and the second microphone was placed next to point SPA, facing upwards, for recording signal d. Signals s and d were recorded independently (not simultaneously).

For capturing auscultated signals, the built-in digital sound-output port of each system was used, if available. For systems with no such built-in ability, the PCB/preamplifier sensing system was used, as described above. Digital sounds were captured using an 8-track ZOOM H4 recorder, situated outside the soundbooth room. All inherent sound effects and sound filters were disabled and the master recording gain was set at 0dB.

Compared systems

The proposed auscultation system was evaluated in terms of the quality of the body sound reaching the ear of the end user, and was further compared to six existing and commercially available systems, including state-of-the art and widely used acoustic or electronic devices, as listed in Table 2.5.

Devices ADC SCOPE and Littmann Cardiology II were evaluated using their tunable diaphragm chestpiece. Device UNICEF only entailed a diaphragm chestpiece. For these three devices, acquisition of the captured body sounds was performed using the PCB microphone, placed into one side of the corresponding earpiece, and was secured with electrical tape. Both sides of the earpiece were then sealed via a 3-step process: i) covered with a thick layer of clay ii) wrapped in multiple layers of acoustic foam, ii) wrapped in multiple layers of cotton cloth. This 3-step process ensured restriction of potential noise leakage through the ends of the earpiece. It is important to highlight here that for such systems were sound travels through a chestpiece, noise leakage is more than likely to occur throughout the full length of the tubing piece. It was outside of the scope of this study to attempt to contain all possible sources of noise leakage, especially those that add to a device's vulnerabilities.

The EKO Core electronic device was toggled to *ON*, to ensure digital acquisition, and the middle volume setting was selected while the diaphragm chestpiece was used for auscultation. Sound acquisition was performed using the PCB Piezotronics condenser microphone, placed into one side of the earpiece, and connected to the recording system. Both sides of the earpiece were then sealed and covered using the 3-step process above. Notice here that the device offers an accompanying phone application for digital sound capturing; and while the bluetooth indicator was flushing, the device was not connected to a phone to ensure uniformity in the recording process. The importance of this experiment is merely to compare the audio signal reaching the ears of the user directly, simulating a scenario of real-time auscultation, and thus, additional computer software was not considered here.

The Littmann 3200 electronic device was set to active mode (non standby mode), the

Table 2.5: List of stethoscope devices and settings

Device	**Type**	**Filter**	**Vol./ max**
ADC SCOPE	Acoustic	Diaphragm piece	N.A.
UNICEF	Acoustic	Diaphragm piece	N.A.
EKO Core	Electronic	Diaphragm piece	3/ 7
Littmann Cardiology II	Acoustic/ Traditional	Diaphragm piece	N.A.
Littmann 3200	Electronic	Extended	3/ 7
Thinklabs One	Electronic	3-4 for Lung Sounds	5/ 10

filter option was set to diaphragm mode, volume set at the middle setting. Sound acquisition was performed using the PCB Piezotronics condenser microphone connected to the recording system, and placed into one side of the earpiece. Both sides of the earpiece were sealed and covered, once again, using the 3-step process above. Due to the restricted automatic shut-off feature, the device had to be set into active mode regularly throughout the duration of the experiments. This device comes with an accompanying computer software for digital sound acquisition; however, the software was not prefered to ensure uniformity in the recording process and a more realistic use-case.

Thinklabs One electronic device does not come with an earpiece, and the auscultation setup is different here. The device offers and audio jack output where the listener/user can connnect custom headphones, providing the capability of directly recording the transmitted sound via an audio cable connected to the recording system. The filter option recommended for lung sounds was used (filter setting 3-4), and the volume was set at the middle point.

Performance metrics

The following metrics were chosen for their i) objective and standardized quality assessment ii) high correlation to human intelligibility scores,[58] and iii) independence to signal amplification or volume variations of the picked up signals.

Normalized Covariance measure (NCM) (section 2.2.2) a speech-based measure accounting for signal audibility at various frequency bands. Though this metric is speech-centric (as are many

quality measures in the literature), it is constructed to account for audibility characteristics of the human ear hence reflecting a general account of improved quality of a signal as perceived by a human listener.

Magnitude squared coherence (MSC) is a statistical index that operates in the spectral domain and examines the relation between two signals. It gives high values for signals that are "close" in coherence to each other.

$$MSC(\omega_k) = 10 \log 10 \frac{1}{K} \sum_{k=1}^{K} \frac{|P_{xy}(\omega_k)|^2}{P_{xy}(\omega_k) P_{xy}(\omega_k)} \tag{2.11}$$

where $P_{xy}(\omega_k)$ is the cross-power spectrum density between signal x and y, with frequency spectrums $X(\omega_k)$ and $Y(\omega_k)$ respectively. For a given frequency band ω_k, the spectral density P_{xy} was estimated by $\hat{P}_{xy} = \sum_{m=1}^{M} X_m(\omega_k) Y_m^*(\omega_k)$ along $m = 1, .., M$ window frames, where M is determined by duration p in (2.14).

The denominator in (2.11) makes the MSC index normalized in [0,1].

Notice that by definition, NCM and MSC are invariant to scalar multiplications of signals x and y; this renders the metrics independent of the volume settings of the individual auscultation systems.

The two metrics, NCM and MSC, were computed over M non-overlapping Hamming windows and were normalized in [0,1]. The duration of the windows varied from short to longer windows: {0.1, 0.5, 1, 2} sec. Both metrics are invariant to amplification of the input signals and obtain values close to 0 when the signals under consideration have low similarity (where the definition similarity is metric-dependent). As a reference, a value of zero would be obtained if one of the compared signal originated from a white Gaussian process. The overal measures of coherence were calculated by averaging the output of the NCM and MSC metrics over all window sizes, yielding quantities $SNR_{lungsound}$ and $SNR_{noisesound}$ that represent the average coherence of the auscultated signal with the true lung sound, and with the net ambient noise respectively. Quantity SNR_{est} represents

the final metric, which accounts for both a) the faithful representation of the truly emitted lung sounds and b) the amount of noise leakage into the auscultated signal:

$$SNR_{est} = SNR_{lungsound} - SNR_{noise} \tag{2.12}$$

$$SNR_{lungsound} = \frac{1}{P_m} \sum_{p=1}^{P_m} Q_p(x, y) \tag{2.13}$$

$$SNR_{noise} = \frac{1}{P_m} \frac{1}{P_m} Q_p(d, y) \tag{2.14}$$

where Q_p is the combined quality metric averaging NCM and MSC values, and p refers to the segmentation windows $\{0.1, 0.5, 1, 2\}$ sec. Signal x corresponds to the true clean breath sound signal driving the chest simulator, y to the measured output signal of a particular auscultation system,

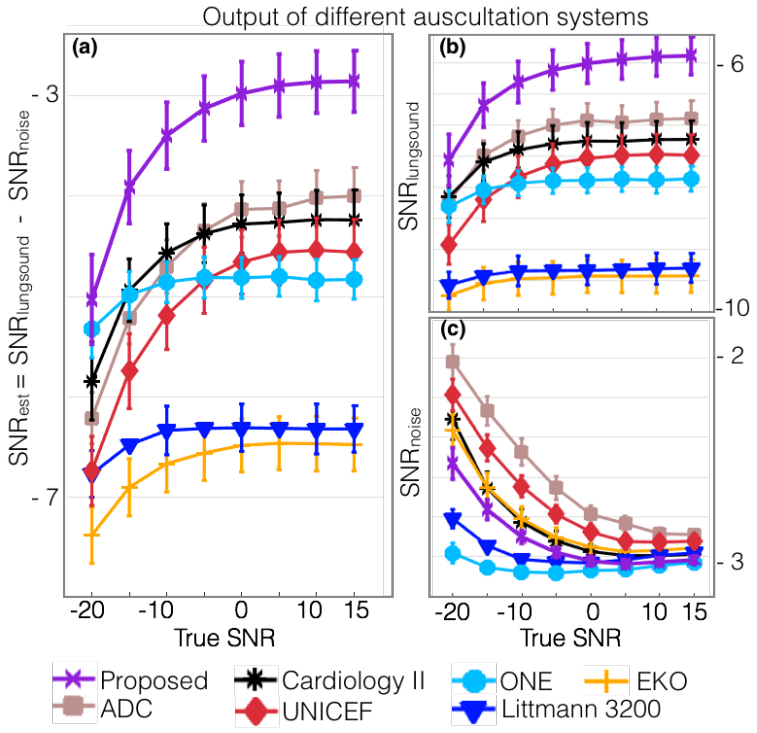

Figure 2.15: Illustration of the sound-preservation ability of different auscultation systems, with varying simulated noise levels. The true SNR is depicted on the x-axis and the estimated SNR is plotted on the y-axis. Main panel (a) depicts results for calculated metric SNR_{est}; panel (b) depicts the $SNR_{lungsound}$ metric, and panel (c) the SNR_{noise} metric. In panels (a-b) high values correspond to high quality of the pick-up signal; in panel (c) high values correspond to maximal noise leakage.

and d to the concurrent net ambient noise during auscultation.

Results:

Results in this section show the performance of the proposed auscultation system when tested within a simulated noisy environment where body sounds where transmitted at Signal to Noise ratios varying from low (-20 dB) to high (15 dB); in this setup a -20 dB SNR condition corresponds to a simulated auscultation in 85 dB SPL noise (on average). The proposed system was also compared with 6 state-of-the-art and commercially available stethoscopes that include a range of acoustic and electronic stethoscope devices, all tested within the same simulated environment, Fig 2.15. Performance curves on the main (left) panel depict i) the similarity between the pick-up signal of a particular system and the true body sound signal driving the chest-simulator ($SNR_{lungsound}$) when accounting for ii) the similarity of the pick-up signal with the ambient noise (SNR_{noise}). The x-axis represents the true SNR (Eq. (2.10)) and the y-axis represents the estimated SNR achieved by the different systems (Eq. (2.14)). High values in the main (left) panell demonstrate increased signal fidelity and reduced noise leakage, while low values depict decreased signal quality and increased noise contamination. The two right panels show the breakdown achieved by the $SNR_{lungsound}$ and SNR_{noise} metrics alone.

The reader is advised not to interpret the results as an absolute comparison measure between the different auscultation devices (see discussion). The main points illustrated by these results are the following: i) some devices perform well in quiet conditions and some are built to withstand noisy environments, where it may be more important to suppress ambient noise leakage than preserving the real signature of the breath sounds (compare the performance of device UNICEF (in red) and device Thinklabs ONE (in cyan)); ii) acoustic stethoscopes perform well in low noise conditions but cannot provide sufficient noise suppression capabilities in noisy settings (Fig. 2.15(c)); iii) electronic stethoscopes can provide advanced filtering to suppress ambient noise but such filtering can also affect the underlying signature of the true breath sounds (device Thinklabs One in cyan,

Littmann 3200 in blue and device EKO core in yellow). The advanced real-time filtering techniques offered by the proposed system (in purple), is shown to effectively suppress unwanted noise leakage, while the sound quality is preserved for all true SNR levels, and the body signal is delivered faithfully with minimum contamination.

Figure 2.16 illustrates the variability in performance curves incurred by the diverse noise database. Variability of performance metric $SNR_{lungsound}$ is shown here at true SNR = -20 dB, for device ADC SCOPE during auscultation of a 10 sec normal lung sound signal. When the simulated noise environment comprised only of High Stationary noise, performance values have low variability within the same auscultation recording Notice that for a long window frame (2 sec) results obtained for High Stationary and Low Stationary noise have both low variability, since a long window tends to

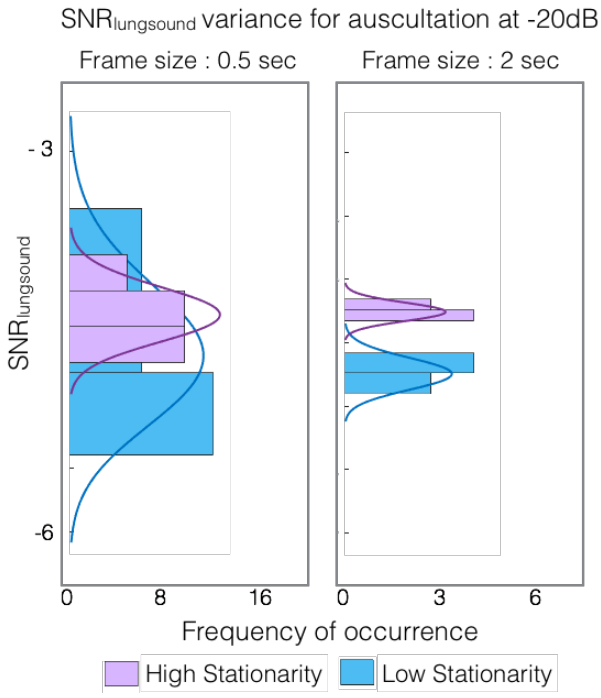

Figure 2.16: Histogram display (bar plots) and fitted gamma distribution curves (solid lines), illustrating the $SNR_{lungsound}$ metric variability for simulated noise environments containing only High and only Low Stationary noise. Variability is high for Low Stationary noise for a smaller window frame (0.5 sec), and is equally low for a higher window frame (2 sec). Results shown here correspond to metric $SNR_{lungsound}$ calculated for device ASC Scope, for true SNR condition of -20 dB, on a 10 sec auscultation of normal breath sounds.

wash out any variability, while for a small size frame (0.5 sec), variability is much greater especially for the Low Stationary case, reflecting the nature of Low Stationary noise.

Conclusion: This work describes a new programmable screening scope, that brings solutions to old and recurring problems of auscultation tools, by entailing an advanced sensing mechanism and dynamic noise suppression design to tackle known shortcomings. It delivers sound signals faithful to the true body sounds, achieving increased pick-up sensitivity and decreased noise leakage. The proposed system was compared with 5 state-of-the art auscultation tools in terms of sound quality and delivery; it's superiority was evaluated using a collection of objective quality measures within a complex simulation environment. Noise leakage and sound alteration effects were evident among all compared systems. And although high noise suppression is desirable, it can incur significant sound alteration to the auscultated sounds. An intermediate solution is thus desirable, where maintenance of the full spectrum of the sound can bring value to the ears of the physician for diagnostic purposes, but also add value to computer-aided auscultation systems (CAAS) that would benefit from a broader representation of body sounds.

The quality performance results presented in this work were based on the standalone capabilities of the included systems, and not on offered supplementary computer programs or phone applications. The objective metrics can reflect the systems' ability to preserve the true emanating body sounds, but should not be used to reflect an expert's ability to form a diagnostic opinion without further exploration.

The proposed system is a robust, powerful and versatile tool, than can be reprogrammed using user-defined algorithms to focus on a variety of applications such as heart and knee-joint auscultation; it can address further aspects of CAAS systems such as suppression subject-specific or application-specific noise; and even add functionalities like automated sound-event detection and decision support frameworks.

Chapter 3

Detecting Respiratory Disease Indicators Using Computerized Methods

Computerized auscultation analyses (CAA) provide a reliable and objective assessment of lung sounds that can inform clinical decisions and may improve case management, especially in resource-poor settings. The challenges in developing such computerized auscultation analysis stem from two main hurdles. Firstly, there is great variability in the literature regarding a reliable description of lung signals and their pathological markers. For instance, adventitious sounds of *wheeze* have been reported to span a wide range of frequencies varying within 100-2500 Hz or 400-1600 Hz; similarly *crackles* have been characterized as sounds with frequency content < 2 kHz or > 500 Hz or within 100-500 Hz.[77,78] Secondly, ambient noise often contaminates the auscultation signal and masks important signature cues, as it often exhibits time-frequency patterns that greatly overlap with characteristic events in lung sounds.[79]

Over the past few decades, few CAA approaches have been proposed to offer solutions

to automated monitoring and diagnosis of lung pathologies. Nonetheless, the proposed approaches remain limited in their applicability, and tend to be confined to laboratory or well-controlled clinical settings or to simulated additive noise conditions.[80–82] These artificial settings greatly oversimplify environments in the field or the Emergency Department, where noisy and raucous clinical conditions incur unpredictable non-additive noise contamination. Few studies have explored analysis and classification techniques for breath sound diagnostics under more realistic clinical settings;[83–87] yet the majority suffers from limited patient evaluation or low protocol versatility. Unfortunately, the applicability of such methods to child auscultation is unknown and expected to be hampered by common pediatric challenges including irregular breathing, motion artifacts, crying or other body sounds that cannot be held back during examination. Finally, most proposed methods offer analysis techniques best suited to only identify context-specific pathological sound patterns.[85–89]

A parallel challenge to the development of fully automated CAA systems is the need for hand-labeled information that can parse the respiratory phases in auscultation signals, identify specific signal instances with pathological markers as well as offer a reference medical interpretation of the auscultation signals. The need for such labeled ground-truth annotations is crucial for the development and training of supervised techniques, which explains why most studies are developed depending on it. Yet, a fully-annotated reference database is unrealistic because: (i) it is an extremely expensive and laborious effort in a large sample size; and (ii) it is not consistent with common medical practices where health care professionals rely on a global listening of the auscultation signal and recurrence of specific patterns indicative of pathologies while ignoring irrelevant information. Requiring an instant-by-instant labeling of hours of auscultation recordings is both unreasonable and impractical.

To tackle these challenges, we introduce a scheme relying on the high quality signal obtained in the previous chapter; (i) it offers a rich feature representation to address the unpredictable nature of adventitious auscultation patterns, and (ii) provides patient-level assessment of pathological status by combining partial signal-level assessments without the need for exhaustively detailed annotations.

For validation and evaluation, we use a large realistic dataset collected in developing countries in non-ideal rural and outpatient clinics. We demonstrate the advantages of the proposed feature extraction against state-of-the-art methods, which are shown here to lack the robustness to perform effectively on a diverse set of adventitious sounds, especially when noise events further mask the signal signatures.

3.1 Feature Extraction

A biomimetic approach was employed to extract meaningful patterns from the enhanced signals, and the acoustic signal was projected onto a high-dimensional space spanning time, frequency as well temporal dynamics and spectral modulations. The analysis followed the model proposed in[90,91] by adapting it to auscultation signals; and is summarized below:

The auscultation signal $s(t)$ was first analyzed through a bank of 128 cochlear filters $h(t;f)$, with 24 channels per octave. These filters were modeled as constant-Q asymmetric band-pass filters and tonotopically arranged with their central frequencies logarithmically spaced. Then, signals were pre-emphasized by a temporal derivative and spectrally sharpened using a first-order difference between adjacent frequency channels, followed by half-way rectification and a short-time integration $\mu(t;\tau)$, with τ=8 ms. The result was an enhanced representation, the auditory spectrogram:

$$y(t,f) = max(\partial_f \partial_t \; s(t) *_f h(t,f), 0) *_t \mu(t;\tau) \tag{3.1}$$

This time-frequency representation was further expanded to extract signal modulations using a multiscale wavelet analysis, akin of processes that take place in the central auditory pathway, particularly at the level of auditory cortex.[91] This analysis yields a rich feature representation that captures intrinsic dependencies and dynamics in the lung sound signals along both time and frequency. This stage is implemented by filtering the auditory spectrogram $y(t,f)$ through a bank of modulation-tuned filters G, selective to specific ranges of modulation in time (rates $\mathfrak{r}$ in Hz) and in frequency

(scales $\mathfrak{s}$ in cycles/octave or c/o):

$$G_+(t, f; \mathfrak{r}, \mathfrak{s}) = A^*(h_r(t; \mathfrak{r}))A(h_s(f; \mathfrak{s})) \tag{3.2a}$$

$$G_-(t, f; \mathfrak{r}, \mathfrak{s}) = A\ (h_r(t; \mathfrak{r}))A(h_s(f; \mathfrak{s})) \tag{3.2b}$$

where A(.) indicates the analytic function, $(.)*$ is the complex conjugate, and $+/-$ indicates upward or downward orientation selectivity in time-frequency space, i.e., detecting upward or downward frequencies sweeping over time: a positive rate corresponds to downward moving energy contents and a negative rate corresponds to upward moving energy contents. The seed functions $h_r(t)$ and $h_s(f)$ were shaped as Gamma and Gabor functions respectively

$$h_r(t) = t^3 e^{-4t} cos(2\pi t), \quad h_s(f) = f^2 e^{1-f^2} \tag{3.3}$$

A filter bank was constructed by dilating the seed function and creating 31 filters of the form $h_r(t; \mathfrak{r}) = \mathfrak{r} h_r(\mathfrak{r} t)$ to capture slow/ fast temporal variations for modulations $\mathfrak{r} = 2^{[1.4:0.22:8]}$; and 21 filters of the form $h_s(f; \mathfrak{s}) = \mathfrak{s} h_s(\mathfrak{s} f)$, to capture narrow/broadband spectral content, with $\mathfrak{s} = 2^{[-5:0.4:3]}$. Each modulation filter output modeled the response of differently-tuned filters, mapping the time waveform onto a high-dimensional space:

$$r_\pm(t, f; \mathfrak{r}, \mathfrak{s}) = y(t, f)\ *_{t,f}\ G_\pm(t, f; \mathfrak{r}, \mathfrak{s}) \tag{3.4}$$

where $*_{t,f}$ corresponds to convolution in time and frequency and $G_\pm$ is the 2D modulation filter response. The final representation was obtained by integrating the response along time, achieving a frequency-rate-scale description:

$$R_\pm(f; \mathfrak{r}, \mathfrak{s}) = \int_t r_\pm(t, f; \mathfrak{r}, \mathfrak{s})\ \delta t \tag{3.5}$$

Note that even though the time axis is integrated in the equation above, details of the temporal changes in the signal are captured along the rate axis $\mathfrak{r}$.

Reduction of feature space dimension

To reduce the size of the feature space, tensor Singular Value Decomposition (SVD) was used. Data was unfolded along each dimension of the SVD space, created by the training data set only. Let R be the feature tensor of order 3 seen above, where the R_- axis is concatenated with the R_+ axis, so that $R \in \mathbb{R}^{d_1 \mathrm{x} d_2 \mathrm{x} d_3}$, where d_1=128 for the frequency axis, d_2= 31x2 = 62 for both $\pm$ rates, and $d_3 = 21$ for scales. When unfolding R along mode (dimension) 1, an order-2 tensor (or matrix) was created, $R^{(1)}$, of dimensions $d_1 \mathrm{x} (d_2 \mathrm{x} d_3)$. Similar order-2 tensors were also created when unfolding along dimension 2 and 3, creating matrices $R^{(2)}$ and $R^{(3)}$. Singular value decompositions were obtained for each of the mode unfoldings $R^{(n)}$, for $n = 1, .., 3$ as:

$$R^{(n)} = U^{(n)} \Sigma^{(n)} V^{(n)\,T} \tag{3.6}$$

For mode-1 unfolding, $\Sigma^{(1)}$ is a diagonal matrix of dimension r, with the nonzero singular values on its diagonal; $r \leq \min\{d_1, (d_2 \mathrm{x} d_3)\}$ is the rank of $R^{(1)}$, i.e. the dimension of the space spanned by the columns or rows of $R^{(1)}$ and $U^{(1)}$ and $V^{(1)\,T}$ are unitary matrices. The singular values in $\Sigma^{(1)}$ are presented ranked, as $\sigma_1^{(1)} > \sigma_2^{(1)} > ... > \sigma_r^{(1)} > 0$. Similar expressions were obtained for mode-2 and mode-3 decomposition. For each $R^{(n)}$, only components capturing up to 99% of the total variance were kept (i.e. $r^{(n)} = \arg\min_x f(x) := \{\sum_{i=1}^{x} \sigma_i^{(n)} \geq 0.99 \,|x = 1, ..., d_n\}$. The final space projection was achieved by tensor-matrix multiplication (mode-n product), significantly reducing the feature dimensions from 128x62x21 to about 5x3x3 (exact dimension can vary depending on the training subset).

3.2 Classification of detected features

The classification of feature vectors into Normal vs. Abnormal was obtained using a soft-margin non-Linear Support Vector Machine (SVM) classifier. Let $\mathbf{x}$ be the matrix comprising of all x_i SVD-projected feature vectors $\in \mathbb{R}^r$, where $r = \prod_{n=1}^{3} r^{(n)}$; and let Φ be a kernel mapping where data is believed to be separable, so that $\Phi(\mathbf{x}) : \mathbf{x} \rightarrow \Phi(\mathbf{x})$, mapping data from $\mathbb{R}^r \rightarrow \mathbb{R}^D$, $D > r$. Given knowledge of data points $\mathbf{x}$, and their true class y, a binary SVM classifier, seeks to learn an optimal hyperplane $\mathbf{w}^T\ \Phi(\mathbf{x})$, $\mathbf{w} \in \mathbb{R}^D$, where

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b \tag{3.7}$$

is the output class participation ($f(\mathrm{x}_i) = \pm 1$) of example x_i; $b = +1 - \mathbf{w}^T\Phi(\mathbf{x})$ for examples in class 1; $b = -1 - \mathbf{w}^T\Phi(\mathbf{x})$ for examples in class -1; and $|\mathbf{w}| = 1$. The optimal hyperplane is found by solving the unconstrained quadratic minimization problem over $\mathbf{w}$:

$$\underset{\mathbf{w}\in\mathbb{R}^D}{min}\ ||\mathbf{w}||^2 + C \sum_{i}^{N} max(0, 1 - y_i\ f(x_i)) \tag{3.8}$$

where N is the number of learning data points and C is a regularization parameter. The second term represents the loss function, where $y_i f(x_i) > 1$ if a data point x_i falls over the correct side of the separating hyperplane margin and $y_i f(x_i) = 1$ if it falls on the margin; finally, $y_i f(x_i) < 1$ if the data point falls on the wrong side of the margin. The optimization problem can also be expressed in its dual form:

$$f(\mathbf{x}) = \sum_{i}^{N} \alpha_i y_i K(x_i, x) + b \tag{3.9}$$

$$\underset{a_i \geq 0}{max} \sum_{i} a_i - \frac{1}{2} \sum_{j,k} a_j a_k y_j y_k k(x_j, x_k) \tag{3.10}$$

subject to $0 \leq a_i \leq C, \forall i$, and $\sum_i a_i y_i = 0$. In the present work, radial-basis kernels (RBF) were used $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) = \exp(-|x(i) - x(j)|^2)$. This way, only the learning of N-dimensional vector $\mathbf{a}$ is needed, avoiding the learning of D-dimensional $\mathbf{w}$ in the primal problem.

3.3 Instrumentation & Implementation

All data and annotations were provided by the Pneumonia Etiology Research for Child Health (PERCH) study.[92] Digital auscultation recordings were acquired from children, ages 1 to 59 months (median age 7 ± 11.43 months), in outpatient or busy clinical settings in Africa (The Gambia, Kenya, South Africa, Zambia) and Asia (Bangladesh, Thailand). In total, 1157 children were enrolled into the digital auscultation study and were classified into one of the two categories: cases, having World Health Organization-defined severe or very severe pneumonia,[93] or age-matched community controls, without clinical pneumonia.

The auscultation protocol called for recordings over 8 body locations (*sites*): four across the child's back, two in the axilla and two on the chest area (Fig. 3.1). To ensure two full breath cycles, at least 7 s of body sounds were obtained per *site*. A commercial digital stethoscope was used for data acquisition (ThinkLabs Inc. ds32a), sampling at 44.1 kHz. An independent Sony-

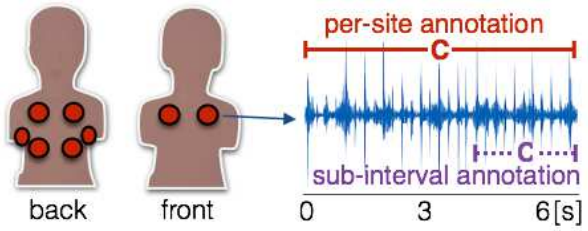

Figure 3.1: Illustration of the 8 auscultation *sites* and the annotation process. A reviewer labeled the depicted *site* as crackles, C, in red/solid line, and then provided an indicative label of a crackling excerpt in purple/dashed line.

ICD-UX71-81 microphone was affixed on the back of the stethoscope, recording concurrent ambient sounds. During examination the infant was seated, laid down or held to the most comfortable position.

Data Collection

Digital auscultation recordings were acquired from children, ages 1 to 59 months (median age 7±11.43 months), in outpatient or busy clinical settings in Africa (The Gambia, Kenya, South Africa, Zambia) and Asia (Bangladesh, Thailand). In total, 1157 children were enrolled into the digital auscultation study and were classified into one of the two categories: cases, having World Health Organization-defined severe or very severe pneumonia,[93] or age-matched community controls, without clinical pneumonia.

The auscultation protocol called for recordings over 8 body locations (*sites*): four across the child's back, two in the axilla and two on the chest area (Fig. 3.1). To ensure two full breath cycles, at least 7 s of body sounds were obtained per *site*. A commercial digital stethoscope was used for data acquisition (ThinkLabs Inc. ds32a), sampling at 44.1 kHz. An independent Sony-ICD-UX71-81 microphone was affixed on the back of the stethoscope, recording concurrent ambient sounds. During examination the infant was seated, laid down or held to the most comfortable position.

Annotations

Nine expert reviewers (pediatricians or pediatric-experienced physicians) were enrolled for the annotation process. For each patient recording, two distinct primary reviewers annotated the 8 sites (*per site* or *site* annotation) as being Normal or Abnormal (Table I), with an accompanying descriptor label: "definite", "probable" or "non-interpretable". A definite label was provided when the reviewer could interpret two or more full breaths with certainty. If only one breath could be

interpreted with certainty or if two or more breaths could be interpreted with uncertainty, then a probable descriptor was given. If no full breath sounds could be distinguished (due to poor sound quality, technical errors, or unrecognizable contamination), a "non-interpretable" label descriptor was assigned.

The above process ensured that every *site* recording was assigned an annotation explaining breath sound findings, along with a confidence indicator for each finding. In case of disagreement between the two primary reviewers, more reviewers listened to the recording to resolve ambiguities, and provided additional labeling as needed (see[94] for details on the annotation process). Finally, within each *per site* label, reviewers were asked to specify a *sub-interval* label containing one segment of arbitrary length that best exemplified the given *per site* label (Fig. 3.1).

Datasets

Based on the *sub-interval* and *per site* labels, two types of data sets were created for the evaluation of this work:

- **Sub-interval set**: including all patients' *sub-interval* recordings of arbitrary length, grouped into Normal and Abnormal (Table 3.1, 1st row).
- **Full patient set**: including all patients' records, grouped as Normal or Abnormal (Table 3.1, 2nd-3rd row).

A few key-observations on the formed data groups: (i) adventitious events may still exist within a normal annotation, as long as their occurrence was not regarded a pathological lung sound; (ii) a *per site* recording was considered abnormal if there was full or partial agreement among reviewers over an abnormal annotation. Full or partial agreement means that a "definite" or "probable" presence of an abnormal sound was agreed by both primary reviewers or by at least two of the total reviewers. Augmenting the data sets to include both full and partial agreement cases ensured the minimization of excluded data, making the study more realistic, but at the expense of

Table 3.1: Available Annotations of Patients' Recordings

Annotation Label	**Abnormal** *(Intervals with wheeze and/or crackles)*	**Normal** *(Intervals without wheeze nor crackles)*
SUB-INTERVAL	annotated clip of arbitrary length found in abnormal *site* recordings of full or partial reviewer agreement	annotated clip of arbitrary length found in normal *site* recordings of full or partial reviewer agreement
PER-SITE (or SITE)	a *site* recording found abnormal by full or partial reviewer agreement	a *site* recording labeled normal by full or partial reviewer agreement
FULL-PATIENT	includes all *site* recordings of a patient if at least one *site* was found abnormal	includes all *site* recordings of a patient when all *sites* were found normal

infusing uncertainty to the classification model; (iii) a patient record labeled as Abnormal (Table 3.1, 3rd row), may contain one or more abnormal *sites* (Table 3.1, 2nd row); (iv) patient records obtaining a "non-interpretable" label or failing to obtain full or partial agreement, were excluded from evaluation.

In total, 62 patients were excluded due to missing annotations, along with 29% of remaining *site* recordings, due to: "non-interpretable" labels, missing audio, recording malfunctions in one of the two microphones, or high disagreement among reviewer labels. The final included data set consisted of more than 250 hours of recorded lung sounds.

Data Processing

All acquired signals were processed using the augmented noise suppression scheme described in Chapter 2, that included clipping correction, heart sound interference suppression, crying elimination, artifact removal and ambient noise suppression.

Timescale of diagnosis

Choosing the timescale (analysis window) over which to perform classification is a nontrivial task. An ideal parsing of the signal would require a window segmentation aligned to the breathing cycle. While this is often the chosen parsing method in studies of limited data,[81,95,96] it is an impractical solution for large datasets recorded in the field: obtaining pre-annotated breath cycles for all subjects

is unrealistic and cannot be automated in a straight-forward manner, especially when considering the irregularity of infant breathing. Alternatively, one could opt for a fixed-size window, which will likely have an impact on the classification outcome. On one end of the spectrum, a very short window will highlight short adventitious events, at the expense of great heterogeneity among training data, especially under noisy conditions. On the other end of the spectrum, a very long window would capture average characteristics of normal vs. abnormal lung sound events but could blend details pertaining to short pathological patterns. We investigated a variety of analysis windows ranging from shorter to longer duration: $W_i \in$ [0.3, ..., 5] s with 50% overlap.

Evaluation of classification results

A closely related issue is the timescale of *evaluating* classification results. The available auscultation dataset contained one annotation per each 7s recording *site*; full-scale, extensive annotations of all sounds of interest were not available and are not a realistic feature, thus, we propose the following algorithmic performance evaluation techniques:

a) *Sub-interval* (used for study comparison in section 3.4): all arbitrary-length *sub-interval annotations* of all available patient records were included in this dataset, grouped into two groups (Normal/Abnormal). A decision for each sub-interval clip was made by the SVM classifier, leading to performance evaluation on the *sub-interval* level;

b) *Full patient* (used for extended evaluation of proposed method in section 3.4): this dataset combined individual frame decisions of each *site* into an overall patient decision. This is not a trivial task, and our approach was designed to be highly sensitive to abnormal occurrences. First, all grouped *site* recordings were split into individual frames of length $W_i \in [0.3, ..., 5]$ s with 50% overlap, and a classifier decision was made at the frame level. Next, a combined decision for each *site* was obtained as follows: a *site* received an abnormal output label if at least (i) 2 consecutive intervals of α duration were found to be abnormal by the classifier or if at least (ii) β% of all overlapping frames were found to be abnormal; (this approach was partially inspired by the annotation protocol

that the medical experts followed, as described above). Finally, a full patient record was assigned an abnormal label if at least one of its *sites* was found to be abnormal; otherwise the patient record was assigned a normal output label. For each time window W_i, parameters α and β were optimized in [0, 2] s and [30, 70] % respectively.

3.4 Findings and Comparison with State of the Art Methods

Findings

After combining the noise suppression scheme with the rich feature analysis and decision integration,

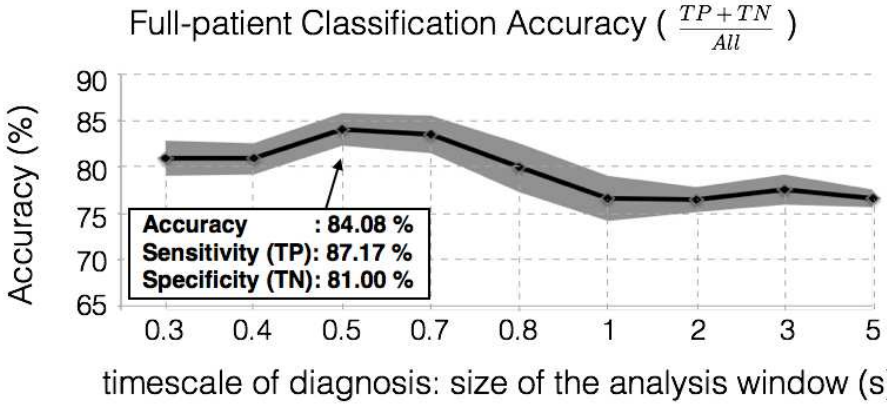

Figure 3.2: Final patient-classification results. Performance was calculated based on the *full-patient decision*; Accuracy = (TP+TN)/All %, where TP: number of True Positives (abnormal patients), TN: number of True Negatives (normal patients), All: total number of patients. Grey shading depicts the standard deviation in patient accuracy among 10 MC runs.

the accuracy of the complete system was assessed for patient-level decisions, using the full-patient evaluation process mentioned above. As outlined earlier, the system performance depends crucially on the choice of analysis window W_i (timescale of diagnosis). Fig. 3.2 shows the system accuracy for different analysis windows. On one hand, large windows > 1 s capture the coarse characteristics of the lung sounds at the expense of the refined detection of adventitious events such as crackle which can be very localized in time and are integrated in these longer time windows. Such coarse analysis yields an accuracy of about 77%. On the other hand, a very short analysis window < 0.5 s can be sensitive to very small or transient changes in the signal hence failing to track sustained patterns

of interest such as wheezes which tend to be very musical in nature and can last few hundreds of milliseconds. Such short windows also yield a smaller drop in accuracy. Overall, it is observed that a balanced time window of about 0.5 s is preferred as it balances the detailed analysis with the tracking of events of interest. Using the recommended 0.5 s, our proposed integrated system yields an overall patient-level accuracy of 84.08% in Fig.3.2. The shaded area shows the standard deviation in accuracy over 10 Monte-Carlo runs.

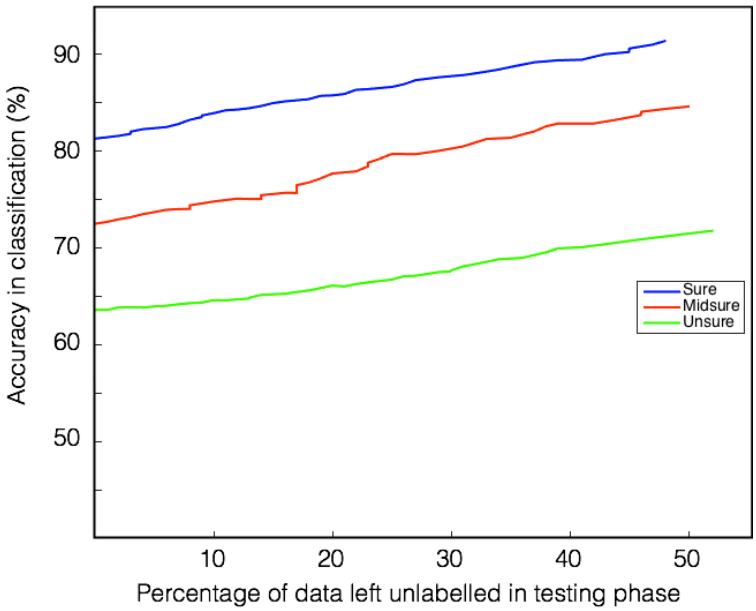

Figure 3.3: Accuracy of classifier with respect to the prercentage of data left unlabelled during the testing phase.

At this point it is worth exploring whether the proposed model will be able to capture the uncertainty in the annotated data. Remember, as discussed earlier, there can be disagreement among the data that the reviewer panel annotated. While the results above reflect only cases where there was agreement between doctors, it is nevertheless worth exploring cases of disagreement. Our efforts are still ongoing but we are able to present some results here. We split the uncertainty level into three groups: Sure, when all doctors agreed on a given annotation, Midsure, when the majority of doctors agreed on a given annotation, and Unsure, when there was no immediate consensus. The model was trained on ALL training examples (Sure, Midsure, Unsure). We altered our gaussian

kernels for the training data to incorporate this uncertainty into the training phase by adjusting the variance of the kernel function according to the type of annotation. Preliminary results shown in Fig. 3.3, demonstrate that our model is able to capture the doctor's opinion, and performs better for cases of less uncertainty. The x axis represents the percentage of data left unlabelled in the testing phase of the classifier (i.e. 100 minus the percentage of data included in the accuracy calculation during the testing phase). The intuition here is that we do not want to classify test cases that fall close to the separating hyperlplane of the classifier. Thus, as we threshold further from the hyperplane, the more data we leave unlabelled. The yaxis represents accuracy (average of Sensitivity and Sensitivity) achieved for the sub-interval classification. From the figure you can see that the group of data annotated with certainty (Sure), i.e. blue curve, achieve the best classification rate. On the other hand, the group with the highest uncertainty during the annotation process (Unsure), i.e. green curve, achieve the lowest accuracy. More work needs to be done to further explore the incorporation of uncertainty into the output labels, but this is a promising result towards the right direction.

Comparison with other methods

The effectiveness of the proposed biomimetic features was furthered explored via a comparison with state of the art methods in the literature. Palaniappan et al. demonstrated the use of the Mel-frequency cepstral coefficients (MFCCs) for capturing spectral characteristics of normal and pathological respiratory sounds.[97] MFCCs are powerful features commonly used in audio signal processing, particularly in speech applications; it is a type of nonlinear cepstral representation calculated on a mel frequency axis, which approximates spectral perception of human listeners:[98] first, the logarithm of the Fourier transform was calculated using the mel scale followed by a cosine transform. One MFCC coefficient was obtained per frequency band, and in total, 13 MFCCs were derived for each data excerpt, averaged over a processing window of 50 ms with 25% overlap. This method is referred to as $MFCC\_P$. In a different study by Jing et al,[99] a new set of discriminating

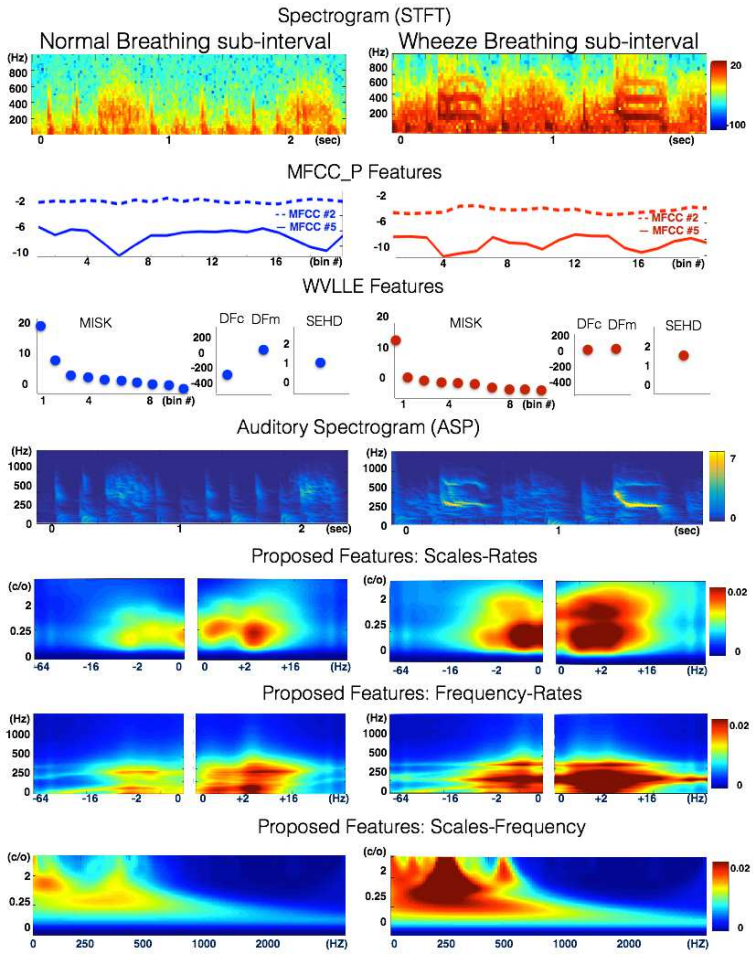

Figure 3.4: Comparison of feature extraction methods for a normal (left) and a wheeze (right) lung sound. Row 1: time-frequency breath characteristics; Row 2-3: binned MFCC coefficients extracted as part of the $MFCC_P$ method, and features MISK, DFc, DFm and SEHD, part of the $WVILLE$ method; Rows 4-7: proposed discriminating features including the auditory spectrogram ASP and the combined spectral and temporal breath dynamics. Notice the high discriminatory nature of the proposed features: the wheezing breath is highlighted with high energy concentration in the Scales-Rates plot ~ 1 c/o, capturing its harmonic structure, and in the Frequency-Rates and Scales-Frequency plots ~ 200 Hz, capturing its pitch. Comparatively, the normal breath exhibits much lower temporal and spectral dynamics.

features was proposed for identifying adventitious events in respiratory sounds, based on spectral and temporal signal characteristics. The features were extracted from a refined spectro-temporal representation, the Gabor time-frequency (Gabor TF) distribution. As the order of the Gabor TF representation increases, it converges to a Wigner-Ville distribution, and we used the latter to extract multiple features from each frequency band, as proposed by the authors: MISK: mean instantaneous

kurtosis, used as feature for adventitious sound localization; DFc and DFm denoting the contrast and minimum value of the calculated discriminating function, used for signal predictability features; and SEHD: sample energy histogram distortion, used as a nonlinear separability criterion for breath discrimination. This method is referred to as $WVILLE$.

For a comparison focused on the effectiveness of the extracted features, we used the data pool created from the *sub-interval* annotations of all subjects in the PERCH database, after full signal enhancement. Recall that the *sub-interval* annotations can be of arbitrary length (with an average duration of 1.8 s in this database). In order to create a relatively uniform database, the intervals were clipped or augmented to 2 s, while intervals shorter than 1 s were discarded.

Fig. 3.4 illustrates the differences of all the feature extraction techniques, as applied on a normal and a wheezing lung sound clip. Row 1 depicts the sound spectrograms calculated on a 30 ms, 50% overlap window simply shown here for reference. Row 2 shows MFCC coefficients #2 and #5 tuned at 75 Hz and 200 Hz respectively, extracted by MFCC_P method. Row 3 shows the WVILLE features: the 10 maximum average instantaneous kurtosis values (MISK); the minimum achieved value of the enclosed discriminating function (DFm) and its center-surround contrast (DFc); and the histogram distortion value (SEHD). Row 4 shows the ASP spectrogram used in the proposed method for extracting the spectro-temporal breath dynamics. Rows 5-7 depict the 3-dimensional Frequency-Rate-Scale space, shown on individual two-dimensional projections. Notice the high discriminatory nature of the proposed set of features: the wheezing breath is highlighted by the presence of strong energy components $\sim$ 1 c/o in the Scales-Rates plot (capturing its harmonic structure), and the energy concentration around 200 Hz along the y-axis of the Frequency-Rates and Scales-Frequency space (capturing its pitch). Compared to the normal breath, the wheezing breath exhibits much higher temporal dynamics as captured by the rates axis.

The RBF SVM classifier was used for all compared methods evaluated on a 10-fold cross validation and 20 Monte Carlo repetitions. Subjects in the training and testing sets were again, mutually exclusive, to avoid classification bias. Recall, that while a normal annotation rules out

Table 3.2: Comparative Classification Results

	Sensitivity (TP)%	Specificity (TN)%	Accuracy%
PROPOSED	86.82 (±0.42)	86.55 (±0.36)	86.67
MFCC_P	91.88 (±0.36)	53.40 (±0.74)	72.64
WVILLE	63.86 (±0.55)	58.47 (±0.60)	61.16

*Performance based on *sub-interval* decision

wheeze or crackle occurrences, the lack of other abnormal sounds such as upper respiratory sounds (URS) or remaining noise cannot be guaranteed, adding real life challenges to the data. Comparative results are shown in Table 3.2, with the accuracy index depicting the average of sensitivity (True Positives Rate) and specificity (True Negatives Rate). The superiority of the proposed feature extraction method was revealed; the rich spectro-temporal space spans intricate details in the lung signal and results in better discriminatory features. Importantly, the proposed features appear to be equally robust in identifying normal and abnormal breath sounds without any bias. In contrast, low accuracy percentages of the $WVILLE$ method are noticeable; the $WVILLE$ features were designed to detect unexpected abnormal patterns within specific breath context, and the feature space seems to lack the ability of separating respiratory-related abnormal sounds from noise-related sounds, signal corruption, or breaths containing possible URS. $MFCC_P$ features were better qualified for identifying abnormal breaths, but when it came to normal segments, both $WVILLE$ and $MFCC_P$ fail to distinguish from noise or other contamination. The $MFCC_P$ and $WVILLE$ methods were previously reported in[97] and[99] to obtain an average accuracy of 77.42% and Area Under the Curve accuracy of 95.60% respectively, in distinguishing normal from pathological lung sounds. However findings of the current work clearly illustrate the inherent difficulty of these feature extraction methods to generalize findings to more realistic or challenging databases and auscultation scenarios.

Chapter 4

Concluding Remarks - Future Work

Over the last decades, there has been an increased interest in computer-aided lung sound analysis. Despite the enthusiasm about possibilities in automated diagnosis, the literature is still shy in tackling real-life challenges. The presented work addresses some of these limitations by proposing a robust discriminative methodology for distinguishing normal and abnormal sounds. Validated on a large-scale realistic dataset, it tackles two aspects crucial in the development of automated auscultation analysis: noise and signal-mapping.

The proposed framework addresses the need for improved lung sound quality by using noise-suppression techniques suitable for auscultation applications. It tackles various noise-sources including ambient noise, signal artifacts, patient-intrinsic maskers (heart-sounds, crying); and explores the use of a rich biomimetic feature-mapping that covers the intricate spectro-temporal details of lung sounds, and yields a notable improvement in distinguishing normal/abnormal events when compared to state-of-the-art systems that tend to fixate on specialized pathologies and global features.

Crucially, the proposed system is further validated on large patient datasets acquired in the field under realistic clinical conditions. The use of such validation data highlights an additional aspect of the analysis; notably the need for full-patient decisions. Previous studies commonly propose methods for localized interpretations on limited pre-segmented breaths; this entails restricted real-life applicability since it requires a pre-segmentation process that is extremely challenging. Instead, our work hopes to take a step towards realistic applicability of computer-aided diagnosis.

A number of challenges remain to be addressed including establishing the association between auscultations and other clinical markers; identifying overlapping non-pathological sounds which can incur significant false positives; and calibrating analysis-windows with respiratory cycles which can benefit the interpretation of the observed patterns. Our continuing efforts focus on the integration of important patient information into the decision making scheme, including body temperature, heart rate, breathing rate, cough occurrence and other. Such augmented patient information supplements the disease detection scheme and can provide an extra layer of robustness to the screening results. Ongoing work further attempts to properly model uncertainty in experts annotations. We have established so far that patient records the yield the highest disagreement amongst doctor experts are also the ones more often confused by our model. Intelligently incorporating such uncertainty and training future models accordingly, will enable computerized methods to learn with confidence from less uncertain patient cases while predicting hard to diagnose patient records.

The following publications have been produced and co-authored as part of this thesis:

Journal Articles & Conference Proceedings

- Park D. et al, Digitally recorded lung sounds in cases and controls compared to standard lung auscultation in the pneumonia etiology research for child health case-control study. 10th Int Symp on Pneumococci and Pneumococcal Diseases, 2016

- McCollum, E. D. et al, Listening panel agreement and characteristics of lung sounds recorded from children aged 159 months enrolled in the PERCH casecontrol study. BMJ Respiratory Research, 4(1), 2017.

- Emmanouilidou D. et al, "Computerized Lung Sound Screening for Pediatric Auscultation in Noisy Field Environments," in IEEE Trans on Biomed Eng, vol. PP, no. 99, pp. 1-1, 2017.

- Emmanouilidou D. et al, "Rich Representation Spaces: Benefits in Digital Auscultation Signal Analysis," 2016 IEEE International Workshop on Signal Processing Systems, Dallas, TX, pp. 69-73, 2016.

- Emmanouilidou D. et al. Adaptive noise suppression of pediatric lung auscultations with real applications to noisy clinical settings in developing countries. IEEE Trans on Biomed Eng. 62(9):2279-88, 2015.

- Ellington L. et al, Developing a reference of normal lung sounds in healthy Peruvian children. Lung Journal, 192:765773, 2014.

- Emmanouilidou D, et al, Characterization of noise contaminations in lung sound recordings. Proceedings of the 2013 35th Annual International Conf Proc IEEE Eng Med Biol Soc: 2551-2554, 2013.

- Emmanouilidou D, et al, A multiresolution analysis for detection of abnormal lung sounds. Conf Proc IEEE Eng Med Biol Soc, 2012:313942, 2012.

Reports & Abstracts

- McCollum E. et al, Digitally recorded lung sounds and mortality among children 1-59 months old with pneumonia, in the Pneumonia Etiology research for Child Health study. Int Amer Thoracic Society, 2017.

- McCollum E. et al, The characteristics and reliability of pediatric digital lung sound examinations in six African and Asian countries participating in the Pneumonia Etiology Research for Child Health (PERCH) project. American Thoracic Society, 2016.

- Emmanouilidou D et al, Programmable electronic stethoscope for improved auscultation analysis. Malone center symposium on engineering in healthcare, Engineering Innovation for Clinical Impact, JHU, 2016.

McCollum E. et al, Diagnosing radiographic lung disease in children with clinical pneumonia using digitally recorded lung sounds: a pneumonia etiology research for child health (perch) sub-study. 10th International Symp on Pneumococci and Pneumococcal Diseases, 2016.

Awards & Intellectual Property

- Finalist award at CIC National Inventors Hall of Fame, DC, USA, 2015.

- Prov. Patents (2) Programmable Electronic Stethoscope, 2015; and Software Algorithm for Denoising of Lung Sound Recordings from a Digital Stethoscope, 2014.

Bibliography

[1] R. L. Wilkins, . Hodgkin, John E. (John Elliott), B. Lopez, and R. L. L. s. Wilkins, *Fundamentals of lung and heart sounds*, 3rd ed. St. Louis, Mo. ; [London] : Mosby, 2004.

[2] (1999) The PERCH (pneumonia etiology research for child health) project. www.jhsph.edu/research/centers-and-institutes/ivac/projects/perch/.

[3] H. N. Wigder, D. R. Johnson, S. Cohan, R. Felde, and R. Colella, "Assessment of lung auscultation by paramedics." *Annals of emergency medicine*, vol. 28, no. 3, pp. 309–12, 1996.

[4] L. H. Brown, J. E. Gough, D. M. Bryan-Berg, and R. C. Hunt, "Assessment of breath sounds during ambulance transport." *Annals of emergency medicine*, vol. 29, no. 2, pp. 228–231, 1997.

[5] N. Al-Naggar, "A new method of lung sounds filtering using modulated least mean squareAdaptive noise cancellation," *Journal of Biomedical Science and Engineering*, vol. 2013, no. September, pp. 869–76, 2013.

[6] M. Molaie, S. Jafari, M. Moradi, J. Sprott, and S. Golpayegani, "A chaotic viewpoint on noise reduction from respiratory sounds," *Biomedical Signal Processing and Control*, vol. 10, pp. 245–9, 2014.

[7] D. Emmanouilidou and M. Elhilali, "Characterization of noise contaminations in lung sound recordings." in *Engineering in Medicine and Biology Society*, vol. 2013, 2013, pp. 2551–4.

[8] D. Emmanouilidou, E. D. McCollum, D. E. Park, and M. Elhilali, "Adaptive Noise Suppression of Pediatric Lung Auscultations with Real Applications to Noisy Clinical Settings in Developing Countries," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 9, pp. 2279–2288, 2015.

[9] G. Nelson, R. Rajamani, and A. Erdman, "Noise control challenges for auscultation on medical evacuation helicopters," *Applied Acoustics*, vol. 80, no. 0, pp. 68 – 78, 2014.

[10] B.-Y. Lu, M.-N. Hsueh, Y.-L. Weng, and S.-H. Tang, "Reduction of the noise in the respiration sound recording by the optimal sampling rate of sound card: The verification by simple filters," *2013 15th International Conference on Advanced Communications Technology (ICACT)*, pp. 148–153, 2013.

[11] Y. Jiao, R. Y. P. Cheung, W. W. Y. Chow, and M. P. C. Mok, "A novel gradient adaptive step size LMS algorithm with dual adaptive filters," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2013, pp. 4803–4806.

[12] F. Ghaderi, H. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 12, pp. 3360–7, Dec 2011.

[13] L. J. Hadjileontiadis, "Lung Sounds: An Advanced Signal Processing Perspective," *Synthesis Lectures on Biomedical Engineering*, vol. 3, no. 1, pp. 1–100, 2008. [Online]. Available: http://www.morganclaypool.com/doi/abs/10.2200/S00127ED1V01Y200811BME009

[14] X. Lu and M. Bahoura, "An integrated automated system for crackles extraction and classification," *Biomedical Signal Processing And Control*, vol. 3, no. 3, pp. 244–54, 2008.

[15] 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. 903–7, 2008.

[16] Z. Li, X. Wu, and M. Du, "A novel method for feature extraction of crackles in lung sound," in *2012 5th International Conference on Biomedical Engineering and Informatics, BMEI 2012*, 2012, pp. 399–402.

[17] L. Zhenzhen and W. Xiaoming, "Wheeze detection using fractional Hilbert transform in the time domain," in *Biomedical Circuits and Systems Conference (BioCAS), 2012 IEEE*, 2012, pp. 316–319.

[18] M. Yamashita, M. Himeshima, and S. Matsunaga, "Robust classification between normal and abnormal lung sounds using adventitious-sound and heart-sound models," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2014, pp. 4418–4422.

[19] S. Emrani and H. Krim, "Wheeze detection and location using spectro-temporal analysis of lung sounds," in *Proceedings - 29th Southern Biomedical Engineering Conference, SBEC 2013*, 2013, pp. 37–38.

[20] H. Pasterkamp, "The R.A.L.E. repository," 2003. [Online]. Available: http://www.rale.ca

[21] J. Racineux, "L'auscultation à l'écoute du poumon ASTRA, CD-Phonopneumogrammes," 1994.

[22] S. Lehrer, *Understanding Lung Sounds*. W.B. Saunders, 2002. [Online]. Available: https://books.google.com/books?id=LURrAAAACAAJ

[23] R. Palaniappan and K. Sundaraj, "Respiratory sound classification using cepstral features and support vector machine," in *2013 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2013*, 2013, pp. 132–136.

[24] M. Bahoura, M. Hubin, and M. Ketata, "Respiratory sounds denoising using wavelet packets," in *Bioelectromagnetism, 1998. Proceedings of the 2nd International Conference on*, Feb 1998, pp. 11–2.

[25] A. Suzuki, C. Sumi, K. Nakayama, and M. Mori, "Real-time adaptive cancelling of ambient noise in lung sound measurement," *Medical and Biological Engineering and Computing*, vol. 33, no. 5, pp. 704–8, 1995.

[26] L. E. Ellington, R. H. Gilman, J. M. Tielsch, M. Steinhoff, D. Figueroa, S. Rodriguez, B. Caffo, B. Tracey, M. Elhilali, J. West, and W. Checkley, "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, no. 1, p. e000506, Jan. 2012.

[27] N. Gavriely, M. Nissan, a. H. Rubin, and D. W. Cugell, "Spectral characteristics of chest wall breath sounds in normal subjects." *Thorax*, vol. 50, no. 12, pp. 1292–300, Dec. 1995.

[28] A. R. A. Sovijärvi, L. P. Malmberg, G. Charbonneau, and J. Vanderschoot, "Characteristics of breath sounds and adventitious respiratory sounds," *European Respiratory Review*, pp. 591–96, 2000.

[29] A. Davignon, P. Rautaharju, E. Boisselle, F. Soumis, M. Mégélas, and A. Choquette, "Normal ecg standards for infants and children," *Pediatric Cardiology*, vol. 1, no. 2, pp. 123–131, Feb 1980. [Online]. Available: https://doi.org/10.1007/BF02083144

[30] 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. 674–84, 2009.

[31] D. Emmanouilidou, K. Patil, J. West, and M. Elhilali, "A multiresolution analysis for detection of abnormal lung sounds," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, Aug 2012, pp. 3139–42.

[32] 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, no. 10, pp. 903–7, Dec. 2008.

[33] L. E. Ellington, D. Emmanouilidou, M. Elhilali, R. H. Gilman, J. M. Tielsch, M. Chavez, J. M. Concha, D. Figueroa, J. West, and W. Checkley, "Developing a reference of normal lung sounds in healthy Peruvian children," *LUNG*, no. in press, 2014.

[34] N. Meslier, G. Charbonneau, and J.-L. Racineux, "Wheezes," *European Respiratory Journal*, vol. 8, no. 11, pp. 1942–8, Nov. 1995.

[35] P. Piirilä and A. Sovijärvi, "Crackles: recording, analysis and clinical significance," *European Respiratory Journal*, vol. 8, no. 12, pp. 2139–48, Dec. 1995.

[36] B. Flietstra, N. Markuzon, A. Vyshedskiy, and R. Murphy, "Automated analysis of crackles in patients with interstitial pulmonary fibrosis," *Pulmonary medicine*, no. 2, pp. 5905–06, 2011.

[37] H. Pasterkamp, M. Montgomery, and W. Wiebicke, "Nomenclature Used by Health Care Describe Breath Sounds in Asthma," *Chest*, vol. 92, no. 2, pp. 346–52, 1987.

[38] F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," *Biomedical engineering*, vol. 58, no. 12, pp. 3360–3367, Dec 2011.

[39] J. Gnitecki, I. Hossain, H. Pasterkamp, and Z. Moussavi, "Qualitative and quantitative evaluation of heart sound reduction from lung sound recordings," *Biomedical engineering*, vol. 52, no. 10, pp. 1788–1792, Oct 2005.

[40] D. Flores-Tapia, Z. M. K. Moussavi, and G. Thomas, "Heart sound cancellation based on multiscale products and linear prediction," *Biomedical engineering*, vol. 54, no. 2, pp. 234–243, Feb 2007.

[41] J. Pesquet, H. Krim, and H. Carfantan, "Time-invariant orthonormal wavelet representations," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1964–1970, 1996.

BIBLIOGRAPHY

[42] P. Basu, D. Rudoy, and P. J. Wolfe, "A nonparametric test for stationarity based on local fourier analysis," *Acoustics*, pp. 3005–3008, 2009.

[43] L. L. LaGasse, A. R. Neal, and B. M. Lester, "Assessment of infant cry: Acoustic cry analysis and parental perception," *Mental Retardation and Developmental Disabilities Research Reviews*, vol. 11, no. 1, pp. 83–93, 2005.

[44] Y. Kheddache and C. Tadj, "Acoustic measures of the cry characteristics of healthy newborns and newborns with pathologies," *J Biomed Sc Eng*, vol. 06, no. 08, pp. 796–804, 2013.

[45] J. L. Goldstein, "An optimum processor theory for the central formation of the pitch of complex tones," *Journal of the Acoustical Society of America*, vol. 54, pp. 1496–1516, 1973.

[46] Johns Hopkins Hospital, K. Arcara, M. Tschudy, M. M. Tschudy, and K. M. Arcara, *The Harriet Lane Handbook: Mobile Medicine Series - Expert Consult*, 19th ed. Philadelphia: Elsevier Mosby, 2011.

[47] G. Prasad, "A Review of Different Approaches of Spectral Subtraction Algorithms for Speech Enhancement," *Current Research in Engineering, Science and Technology Journals*, vol. 01, no. 02, pp. 57–64, 2013.

[48] Philipos C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed. Boca Raton, FL: CRC Press, 2013.

[49] P. Vary, "Noise suppression by spectral magnitude estimation-mechanism and theoretical limits," *Signal Processing*, vol. 8, pp. 387–400, 1985.

[50] A. R. A. Sovijärvi, J. Vanderschoot, and J. E. Earis, "Standardization of computerized respiratory sound analysis," *European Respiratory Review*, vol. 10, no. 77, p. 585, 2000.

[51] J. Beh and H. Ko, "Spectral Subtraction Using Spectral Harmonics for Robust Speech Recognition in Car Environments," in *International Conference of Computational Science*, 2003, pp. 1109–16.

BIBLIOGRAPHY

[52] W.H.O. (2006) Pocket book of hospital care for children: guidelines for the management of common illnesses with limited resources.

[53] L. L. Schumaker, *Spline functions : basic theory*, ser. Pure and applied mathematics : a Wiley-Interscience series of texts, monographs, and tracts. New York, Chichester, Brisbane: J. Wiley & Sons, 1981.

[54] S. Reichert, R. Gass, C. Brandt, and E. Andrès, "Analysis of respiratory sounds: state of the art." *Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine*, vol. 2, pp. 45–58, Jan. 2008.

[55] 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." *Respiratory medicine*, vol. 105, no. 9, pp. 1396–403, Sep. 2011.

[56] E. Vincent, R. Gribonval, and C. Févotte, "Performance Measurement in Blind Audio Source Separation," *IEEE Transactions on Speech and Audio Processing*, vol. 14, no. 4, p. 1462, 2006.

[57] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality*, 1st ed. Englewood Cliffs, NJ: Prentice Hall, 1998.

[58] J. Ma, Y. Hu, and P. C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions." *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3387–405, may 2009.

[59] A. S3.5-1997, "American National Standard Methods for Calculation of the Speech Intelligibility Index," New York, 1997.

[60] J. M. Kates and K. H. Arehart, "Coherence and the speech intelligibility index," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, p. 2224, 2005.

[61] I. Hossain and Z. Moussavi, "An overview of heart-noise reduction of lung sound using wavelet transform based filter," in *Engineering in Medicine and Biology Society. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 1, Sept 2003, pp. 458–61.

[62] G. Chang and Y. Lai, "Performance evaluation and enhancement of lung sound recognition system in two real noisy environments," *Computer Methods and Programs in Biomedicine*, vol. 97, no. 2, pp. 141–50, 2010.

[63] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 27, no. 2, pp. 113–20, Apr 1979.

[64] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 4, Apr 1979, pp. 208–11.

[65] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (nss), hidden markov models and the projection, for robust speech recognition in cars," *Speech Communication*, vol. 11, no. 2-3, pp. 215–28, 1992.

[66] L. Singh and S. Sridharan, "Speech Enhancement using Critical Band Spectral Subtraction," in *Proceedings of International Conference on Spoken Language Processing, Sydney, Australia*, 1979, pp. 2827–30.

[67] S. Patel, T. Callahan, and M. e. a. Callahan, "An adaptive noise reduction stethoscope for auscultation in high noise environments," *J. Acoust. Soc. Am.*, vol. 103, pp. 2483–91, may 1998.

[68] S. Kuo and D. Morgan, "Active noise control: a tutorial review," *Proceedings of the IEEE*, vol. 87, no. 6, pp. 943–73, Jun 1999.

[69] S. Haykin, *Adaptive Filter Theory (3rd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[70] J. Valin and I. Collings, "Interference-normalized least mean square algorithm," *Signal Processing Letters, IEEE*, vol. 14, no. 12, pp. 988–91, Dec 2007.

[71] E. V. Kuhn, J. E. Kolodziej, and R. Seara, "Stochastic modeling of the nlms algorithm for complex gaussian input data and nonstationary environment," *Digital Signal Processing*, vol. 30, pp. 55 – 66, 2014.

[72] C. Van Sligtenhorst, D. S. Cronin, and G. Wayne Brodland, "High strain rate compressive properties of bovine muscle tissue determined using a split Hopkinson bar apparatus," *Journal of Biomechanics*, vol. 39, no. 10, pp. 1852–1858, 2006.

[73] M. L. Fackler and J. A. Malinowski, "Ordnance gelatin for ballistic studies. Detrimental effect of excess heat used in gelatin preparation." *The American journal of forensic medicine and pathology*, vol. 9, no. 3, pp. 218–219, sep 1988.

[74] J. A. S. McCann, *Nursing know-how: Evaluating heart & breath sounds.* Phialdelphia: Lippincott Williams & Wilkins, 2009, iD: 731183533.

[75] "The bbc sound effects library original series," *http://www.soundideas.com*, May 2006.

[76] A. P. Varga, H. J. M. Steeneken, M. Tomlinson, and D. Jones, "The noisex-92 study on the effect of additive noise on automatic speech recognition," *Tech.Rep., Speech Research Unit, Defense Research Agency, Malvern, U.K.*, 1992.

[77] S. Reichert, R. Gass, C. Brandt, and E. Andres, "Analysis of respiratory sounds: state of the art." *Clinical medicine Circulatory respiratory and pulmonary medicine*, vol. 2, pp. 45–58, 2008.

[78] B. Flietstra, N. Markuzon, A. Vyshedskiy, and R. Murphy, "Automated analysis of crackles in patients with interstitial pulmonary fibrosis," *Pulmonary medicine*, no. 2, p. 1, 2011.

[79] D. Emmanouilidou and M. Elhilali, "Characterization of noise contaminations in lung sound recordings," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 2551–2554.

[80] N. Q. Al-Naggar, "A new method of lung sounds filtering using modulated least mean squareadaptive noise cancellation," *Journal of Biomedical Science and Engineering*, vol. 6, pp. 869–876, 2013.

[81] 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," *Journal of Asthma*, vol. 45, no. 10, pp. 903–907, 01/01 2008.

[82] J. Li and Y. Hong, "Wheeze detection algorithm based on spectrogram analysis," in *2015 8th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 1, 2015, pp. 318–322.

[83] D. Emmanouilidou., E. D. McCollum, D. E. Park, and M. Elhilali, "Adaptive noise suppression of pediatric lung auscultations with real applications to noisy clinical settings in developing countries," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 9, pp. 2279–2288, Sept 2015.

[84] S. B. Patel, T. F. Callahan, M. G. Callahan, J. T. Jones, G. P. Graber, K. S. Foster, K. Glifort, and G. R. Wodicka, "An adaptive noise reduction stethoscope for auscultation in high noise environments," *The Journal of the Acoustical Society of America*, vol. 103, no. 5 Pt 1, pp. 2483–2491, May 1998.

[85] A. Poreva, Y. Karplyuk, A. Makarenkova, and A. Makarenkov, "Application of bispectrum analysis to lung sounds in patients with the chronic obstructive lung disease," in *Electronics and Nanotechnology (ELNANO), 2014 IEEE 34th International Conference on*, 2014, pp. 306–309.

[86] M. Lozano, J. A. Fiz, and R. Jan, "Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous

frequency," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 2, pp. 486–497, 2016.

[87] G. Nelson and R. Rajamani, "Accelerometer-based acoustic control: Enabling auscultation on a black hawk helicopter," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 2, pp. 994–1003, April 2017.

[88] R. M. Rady, I. M. E. Akkary, A. N. Haroun, N. A. E. Fasseh, and M. M. Azmy, "Respiratory wheeze sound analysis using digital signal processing techniques," in *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*, 2015, pp. 162–165.

[89] N. Nakamura, M. Yamashita, and S. Matsunaga, "Detection of patients considering observation frequency of continuous and discontinuous adventitious sounds in lung sounds," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3457–3460, iD: 1.

[90] D. Emmanouilidou, K. Patil, J. West, and M. Elhilali, "A multiresolution analysis for detection of abnormal lung sounds," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2012, pp. 3139–3142.

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

[92] O. S. Levine, K. L. O'Brien, M. Deloria-Knoll, D. R. Murdoch, D. R. Feikin, A. N. DeLuca, A. J. Driscoll, H. C. Baggett, W. A. Brooks, S. R. Howie, K. L. Kotloff, S. A. Madhi, S. A. Maloney, S. Sow, D. M. Thea, and J. A. Scott, "The pneumonia etiology research for child health project: a 21st century childhood pneumonia etiology study," *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, vol. 54 Suppl 2, pp. S93–101, Apr 2012.

[93] W. H. Organization, *Pocket book of hospital care for children: guidelines for the management of common illnesses with limited resources*. Geneva, Switzerland: WHO press, 2005.

[94] E. D. McCollum, D. E. Park, N. Watson, W. C. Buck, C. Bunthi, A. Devendra, B. E. Ebruke, M. Elhilali, D. Emmanouilidou, A. J. Garcia-Prats, L. N. Githinji, M. L. Hossain, D. P. Moore, A. Mudau, J. M. Mulindwa, D. Olson, J. O. Awori, W. P. Vandepitte, C. Verwey, J. E. West, K. L. O'Brien, D. Feikin, and L. Hammitt, *The Characteristics and Reliability of Pediatric Digital Lung Sound Examinations in Six African and Asian Countries Participating in the Pneumonia Etiology Research for Child Health (PERCH) Project*. American Thoracic Society, 2016, p. A3043.

[95] 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, jul 2008.

[96] L. Zhenzhen, W. Xiaoming, and D. Minghui, "A novel method for feature extraction of crackles in lung sound," in *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*, 2012, pp. 399–402.

[97] R. Palaniappan and K. Sundaraj, "Respiratory sound classification using cepstral features and support vector machine," in *Intelligent Computational Systems (RAICS), 2013 IEEE Recent Advances in*, 2013, pp. 132–136.

[98] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 4, pp. 357–366, 1980.

[99] F. Jin, F. Sattar, and D. Y. T. Goh, "New approaches for spectro-temporal feature extraction with applications to respiratory sound classification." *Neurocomputing*, vol. 123, pp. 362–371, 2014.

Curriculum Vitae

Dimitra Emmanouilidou was born in Greece in 1984. She received her BS degree in Computer Science from University of Crete, Greece, in 2007 and her MSc in Bioinformatics & Technology from University of Crete, Greece, in 2010. She enrolled in the Electrical & Engineering Ph.D. program at Johns Hopkins in 2011. She is the co-inventor of a smart screening tool for detecting respiratory diseases, powered by advanced signal processing and artificial intelligence. Her focus and interests are in improving existing sensing technologies, extracting meaningful patterns from data and creating intelligent decision-support systems.