Validation of Auscultation Technologies using Objective and Clinical Comparisons

Stephanie Graceffo[1], Amyna Husain[2], Salahuddin Ahmed[3], Eric D. McCollum[4], and Mounya Elhilali[1]

[1]Johns Hopkins University, Department of Electrical and Computer Engineering
[2]Johns Hopkins University, Department of Pediatrics, Pediatric Emergency Medicine
[3]Projahnmo Research Foundation, Dhaka, Bangladesh
[4]Johns Hopkins School of Medicine, Global Program of Pediatric Respiratory Sciences, Eudowood Division of Pediatric Respiratory Sciences, Department of Pediatrics

***Abstract*— Technology is rapidly changing the health care industry. As new systems and devices are developed, validating their effectiveness in practice is not trivial, yet it is essential for assessing their technical and clinical capabilities. Digital auscultations are new technologies that are changing the landscape of diagnosis of lung and heart sounds and revamping the centuries old original design of the stethoscope. Here, we propose a methodology to validate a newly developed digital stethoscope, and compare its effectiveness against a market-accepted device, using a combination of signal properties and clinical assessments. Data from 100 pediatric patients is collected using both devices side by side in two clinical sites. Using the proposed methodology, we objectively compare the technical performance of the two devices, and identify clinical situations where performance of the two devices differs. The proposed methodology offers a general approach to verify a new digital auscultation device as clinically-viable; while highlighting the important consideration for clinical conditions in performing these evaluations.**

I. Introduction

The stethoscope is a well-established tool for diagnosing lung and heart diseases. While this universal tool is found in clinics across the globe, it has several limitations impairing the physician's ability to accurately assess cardiac or pulmonary sounds, including chest-piece placement sensitivity and signal degradation due to background noise. Advancements in this technology, however, have been proposed to address some of these limitations, especially in unusual clinical settings where healthcare access can be limited. Eko Devices developed a device that uses amplification and Active Noise Cancellation then wirelessly streams the collected data to a smartphone app for review by a trained physician [1], [6]. MIT developed a low-cost diagnostic tool that incorporates lung sounds with other data and provides diagnostic feedback to the healthcare workers [17]. Our team at Johns Hopkins has also developed a digital stethoscope that uses a microphone array, rather than a single point microphone, and an adaptive noise cancellation algorithm, making chest-piece placement on the patient less critical and mitigating background noise in noisy clinics [18].

As new technology becomes available, it must be validated against commercial standards. Validation methods should include a balance between objective comparisons using signal processing and analytical methods; and clinical evaluations that consider use cases and medical outlooks. Stethoscopes deliver to the physician's ear the acoustic energy from a body organ, particularly the lungs, which reflects presence of any abnormality in the respiratory system. Validating a new technology that provides access to this signal involves assessing the spectral and temporal characteristics of the auscultation signal delivered by the device. An objective comparison against existing market technologies helps identify trends as well as similarities and differences among various systems. Statistical similarities between the systems' characteristics indicate that the new technology works as well as the current technology. Differences in these statistics, however, may not necessarily invalidate one technology or the other, but instead, they help define unique use cases or delineate potential limitations of some devices. For instance, stethoscopes typically shape the spectral content of a lung sound either through their design (e.g. tubing, transducers) or in software in order to control the amount of leakage of ambient noises into the actual body sound. Use of overly-aggressive filtering of spectral content may benefit users by blocking presence of background disturbances; but could also limit the effectiveness of diagnosing abnormal auscultations such as crackles or high-pitched wheezes, which sometimes can masquerade as noise. This balance of high quality sound with effective diagnosis can often be difficult to gauge in a laboratory setting under controlled conditions and requires careful consideration of use case and clinical environment in order to properly evaluate utility and impact of an auscultation technology.

In the present work, we propose a validation methodology for the JHUscope, a new system developed by our group that contains hardware and software improvements over many existing stethoscopes [18]. This device contains a microphone array consisting of five microphones that face inward towards the patient, making it easier to pick up lung sounds even if the stethoscope is not ideally placed on the body, and a sixth microphone facing outward, that collects noise from the environment. Pathological conditions create abnormal signal patterns, and it is essential that any noise-filtering leave these critical pieces of information intact. The team

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

developed onboard noise-cancellation algorithm that uses active and adaptive acoustics. The algorithm analyzes lung and ambient sounds recorded from the device microphones and performs an adaptive noise cancellation to deliver a cleaner lung sound to the physician's ear. In order to validate the efficacy of this device, we perform analytical and clinical analyses of its auscultation recordings, and compare them against a commercially and clinically-approved device by performing signal comparisons and evaluating use cases. The validation of the JHUscope is performed against the Thinklabs One digital stethoscope. This device was designed to eliminate the hollow tube styling that has been around since the stethoscopes invention in the early 1800's and is considered one of the most powerful stethoscope on the market. It offers 100x amplification, allows users to adjust the range of frequencies heard, and has three filter settings as well as ten volume levels, making it versatile for virtually any user [2].

While we explore this comparison in the context of our new device, the JHUscope, the methodologies outlined here can be extended to a variety of technological devices. The remainder of this paper is organized as follows: *Methods* describes the pre-processing steps, the spectral and spectro-temporal feature computation, and the datasets used in this exploration. *Results* shows the results of our validation procedure, the similarities and differences in the spectra and the features. *Conclusion* summarizes our findings and discusses potential future work.

II. Methods

A. Data

To assess our validation method in a range of clinical challenges, auscultation signals were collected from two clinical sites: 1) the Pediatric Emergency Department (PED) at Johns Hopkins University Hospital and 2) small, rural community clinics in Bangladesh under the Projahnmo Study Group, in the northeastern district of Sylhet in Bangladesh [18]. The hospital data consists of 25 pediatric patients with a mean age of 6.02 years ($\pm$4.05) and was collected in a relatively quiet, controlled environment (i.e., in an exam room with the a closed door). The rural clinic data consists of 75 pediatric patients and was collected in a much noisier environment: local clinics, polluted with background sounds including crying babies, people talking, generators, passing mopeds, and nearby markets [7]. At each clinical site, auscultation signals were collected using the two stethoscopes in a back-to-back procedure from four thoracic positions [9]. All signals were recorded using a ZOOM H4NPro recorder with a 4 GB SD memory card. Whenever possible, auscultations recorded using the JHUscope and Thinklabs devices were augmented with additional clinical protocols (X-rays, blood work), as well as diagnosis using a traditional acoustic stethoscope to note presence of abnormal breathing patterns (wheezes, crackles, etc).

B. Pre-Processing

Signals were originally recorded at 44.1 kHz then downsampled to 8 kHz. For further processing and extraction of characteristic features, all signals were further downsampled to 2 kHz using an anti-aliasing 4-th order Butterworth filter, then normalized to have zero-mean and unit-variance. Losing crucial information during this pre-processing stage was not a concern, as 90% of the total signal energy was concentrated in frequencies below 1 kHz [8]. The signals were the segmented into 500 ms time fragments using a rectangular window with 50% overlap. Any segment whose average amplitude was less than 20% of the signal's average amplitude was classified as "silent" and excluded from the analysis. [14], [13], [10].

C. Feature Analysis

A total of eight spectral and spectro-temporal features were extracted from each recording segment. The spectral features capture frequency content information while the temporal features capture dynamic changes or modulations in the frequency content of the lung sounds. Figure 1 shows the average power spectrum for the JHU data and generated as follows: the STFT was computed using 500 ms windows for each recording in the JHU dataset. The power spectrum for each STFT was computed then averaged together to yield one representative spectrum. The spectral features are annotated in Figure 1 and described below.

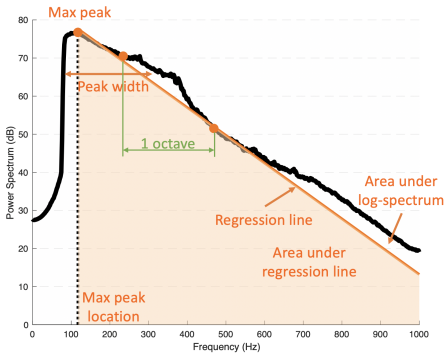

Fig. 1: Annotation of spectral features. Note: The Peak Max, Peak Frequency, and Peak Width are shown for reference but are not computed on the Power Spectrum in dB.

Max Peak - The maximum value in the power spectrum.
Max Peak Location - Frequency of the maximum value (max peak) in the power spectrum.
Peak Width - The peak width of the smoothed power spectrum measured at 75% of the maximum peak [8].
Spectrum Slope - The slope of the linear regression line, fit to the power spectrum in logarithmic axes. The spectrum slope is measured in dB/octave, where an octave represents the interval needed to double the frequency. The power spectrum, plotted in dB, is defined as

$$10 * log(P/P_{min})$$

where $P_{min} = 5 \times 10^{-5}$. [8], [3].

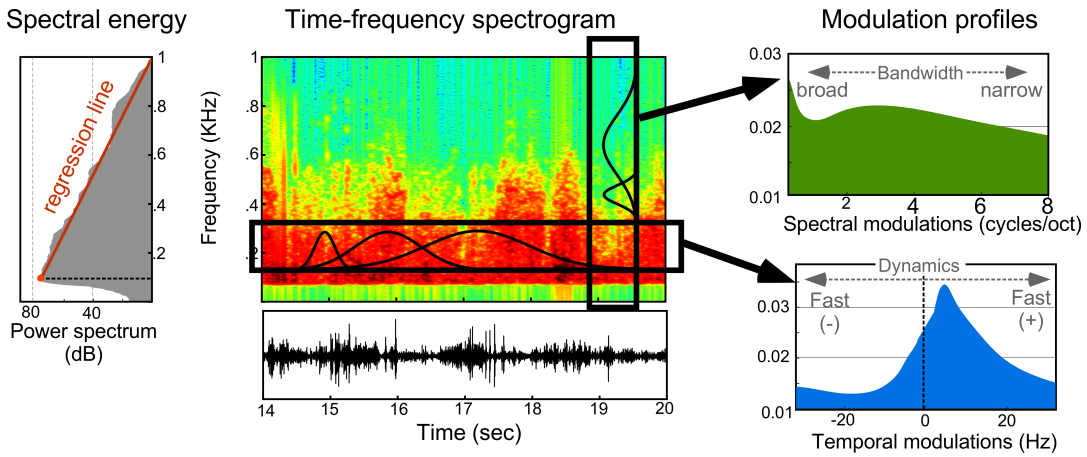

Fig. 2: Schematic representation of Bandwidth and Dynamics.

Power of Regression Line - The power of the area under the regression line [8].

Power Ratio - The ratio of power between the area under the logarithmic spectrum, $E_{spectrum}$ and the area under the linear regression line $E_{regression}$. It is defined as

$$PR = 1 - |1 - E_{spectrum}/E_{regression}|$$

These areas are computed using trapezoidal integration. A power ratio value close to 1 indicates that the logarithmic spectrum closely follows the regression line [8], [3].

Spectral Modulations (bandwidth) - Spectral modulation features estimate how broad or narrow the spectral profile is. These bandwidth attributes reflect how contents vary along frequency and reflect presence of broad or narrow spectral components throughout the Fourier representation of the signal. These features highlight presence of transients (e.g. abnormal crackling lung sounds or sharp broad noises); they also pickup presence of musical sounds such as wheezing in the lung or possible ambient voices. Spectral modulations were calculated from a time-frequency auditory spectrogram following the procedure in Chi et al. [4]. The auditory spectrogram was filtered using 31 Gabor-shape seed filters, logarithmically spaced and varying from wideband to narrowband: 0-8 cycles/octave (c/o) [8], [4], [19]. The response produced for each bandwidth and time index was averaged over time to produce a bandwidth profile. Low bandwidth values correspond to a smooth spectral profile with peaks that spread over more than 1 octave. High bandwidth values correspond to a peaky spectral profile with more than one peaks to troughs within one octave. Figure 2-right panel (top) shows a typical profile of spectral modulations obtained from a lung sound.

Temporal Modulations (dynamics) - Dynamics capture how fast or slow the frequency contents change with time and in which phase (direction), positive or negative. The temporal modulations were calculated from the auditory spectrogram using 23 exponential filters constructed from varying velocities ranging from 0 to 32 Hz for both the positive and negative directions [8], [4], [19]. The direction of the dynamics (positive versus negative) reflects the phase of the modulations as upward or downward moving. Dynamics were computed for each frequency band of the spectrogram and the results were averaged to produce one dynamics profile. Figure 2-right panel (bottom) shows the dynamics derived from an auscultation signal and shows a clear peak at slower temporal rates, with a clear bias to positive modulations.

The feature analysis was divided into two experiments: the first experiment used the quiet hospital data and the second experiment used the noisy clinic data. In each experiment, the Short-Term time-frequency spectrogram was calculated for each patient. During the short-term spectrogram computation, windows determined to be silent segments and therefore non-informative were discarded. Each feature was then computed from the spectrogram to give one value per feature per patient. To quantify the statistical similarities between the two devices, we ran a t-test (results in Tables I and II) and a paired t-test (results not shown here as outcomes were quantitatively and quantitatively similar to those obtained with an unpaired statistical test). For each device, we computed each feature's mean by averaging the feature values across all patients, as well as the difference between the means (results in Tables I and II).

III. Results

Using spectral and spectro-temporal features, we compared auscultation signals obtained from the JHUscope versus the Thinklabs One. Table I shows the side-by-side comparison of all features from patient data obtained in the hospital setting, along with their statistical comparison. As the results indicate, many spectrotemporal features are virtually identical when comparing the two devices. Notably, dynamics, bandwidths as well as power ratio, power regression line, spectral slope, and peak width are all statistically similar. However, the peak max and peak frequency reveal a statistical difference between both devices. Looking closely at the average spectral profile of both devices from data collected in this hospital setting (Figure 3), we note that Thinklabs device has a higher peak, slightly lower peak frequency, and a narrower peak width. This slight deviation

Feature	p-value	JHUscope Mean	Thinklabs Mean	Absolute Difference Between Means
Peak Frequency	<0.001	121 Hz	95.0 Hz	25.7 Hz
Peak Max	<0.001	2.90×10^3	4.93×10^3	2.03×10^3
Peak Width	0.176	7.31 Hz	6.49 Hz	0.825 Hz
Spectral Slope	0.590	-0.155 dB/oct	-0.153 dB/oct	0.001 dB/oct
Power Regression Line	0.779	1.20×10^3	1.18×10^3	24.8
Power Ratio	0.435	-66.41	-39.45	26.95
Bandwidths	0.708	1.36 c/o	1.36 c/o	0.004 c/o
Dynamics (Positive)	0.640	8.44 Hz	8.45 Hz	0.010 Hz
Dynamics (Negative)	0.876	5.61 Hz	5.63 Hz	0.021 Hz

TABLE I: Hospital Data (clean): Average spectral and spectrotemporal feature values and statistics

Feature	p-value	JHUscope Mean	Thinklabs Mean	Absolute Difference Between Means
Peak Frequency	0.030	122 Hz	132 Hz	9.54 Hz
Peak Max	<0.001	3.65×10^3	6.78×10^3	1.64×10^3
Peak Width	0.060	7.20 Hz	6.29 Hz	0.914 Hz
Spectral Slope	<0.001	-0.152 dB/oct	-0.127 dB/oct	0.025 dB/oct
Power Regression Line	<0.001	1.29×10^3	2.26×10^3	971
Power Ratio	<0.001	-35.68	-23.25	12.4
Bandwidth	<0.001	1.36 c/o	1.42 c/o	0.061 c/o
Dynamics (Positive)	<0.001	8.39 Hz	8.25 Hz	0.140 Hz
Dynamics (Negative)	<0.001	5.64 Hz	5.81 Hz	0.173 Hz

TABLE II: Bangladesh Clinic (noisy): Average spectra and spectrotemporal feature values and statistics

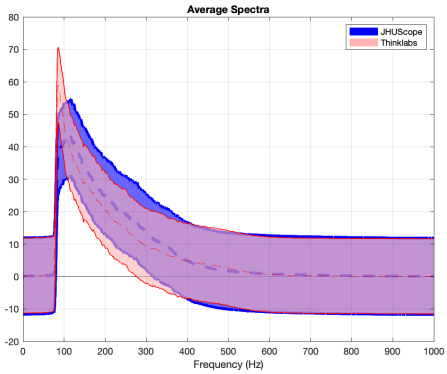

Fig. 3: Spectra of JHUscope and Thinklabs One, averaged over all patients from the quiet hospital data

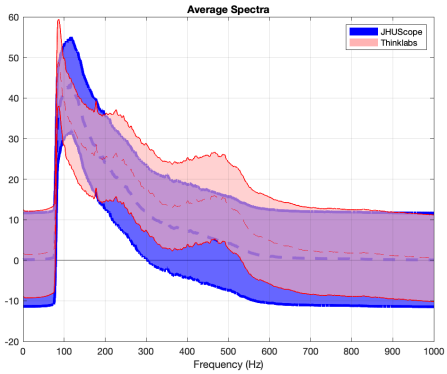

Fig. 4: Spectra of JHUscope and Thinklabs One, averaged over all patients from the rural clinic data

from the JHUscope is concentrated mostly near the area of peak energy in the spectrum (as indicated by the three features: peak max, peak frequency, and peak width) and likely reflects the filtering profile of Thinklabs devices having a slightly more assertive filtering property. Nonetheless, this distinction between the JHUscope and Thinklabs near peak energies in the 95-120 Hz region is unlikely to have a major impact on the perception of lung sounds. Following suggested guidelines for lung sound processing [16], respiratory sounds are mostly perceptible above 100 Hz, with most abnormal signatures (crackles, wheezing, rhonchus) also manifesting above 100 Hz or 300 Hz in some cases [5], [15], [12], [11]. Overall, the average profiles indicate a close correspondence between sounds obtained using the JHUscope and Thinklabs One in the quiet hospital setting.

In contrast, we performed the same analysis comparing signals obtained side by side from the two devices in rural clinics. Table II shows the average value of different features along with a statistical comparison of the two signals. The table clearly indicates a statistical difference between almost all characteristics obtained from the JHUscope and Thinklabs One in the noisy clinic data, the one exception being the peak width. Looking closely at the average spectra from both signals (Figure 4), it is clear that the Thinklabs device reflects a great deal of noise that is prevalent in this clinical setting and tend to manifest itself in higher frequency regions between 200 and 500 Hz. The JHUscope, however, is less affected by ambient noise. This noise leakage in the Thinklabs data results in a shallower spectral roll-off yielding a smaller spectral slope.

Overall, the comparison of signals obtained in the clinical setting suggests that presence of noise yields to substantial difference in the auscultation signal obtained from both devices. We therefore take a closer look at individual patient cases to evaluate whether this difference may have diagnostic implications.

The spectrograms in Figure 5 show clinical situations where important lung sound information identified by the JHUscope is masked by the noise in the Thinklabs device. The top row shows the spectrograms from the JHUscope, and the bottom row shows the spectrograms from the Thinklabs device. Spectrogram pair 5a shows spectrograms for a patient where the JHUscope captured lung sounds containing periodic bursts of low-frequency energy but the Thinklabs did not. The frequency content in the Thinklabs device spikes to 1000 Hz periodically, masking the low-frequency energy bursts. Pair 5b shows a similar pattern to the previous pair: the low-frequency energy bursts can be identified in the JHUscope, but not in the Thinklabs device. The noise does not spike to 1000 Hz as in the previous pair, but the noise blurs the respiratory cycles together, making the low-frequency content virtually inaudible. In the JHUscope audio signals for 5a and 5b, there were audible grunting sounds, which is often indicative of a respiratory infection; these noises were not audible in the corresponding Thinklabs audio signals. The spectrogram pair in 5c shows that the Thinklabs device missed some of the higher frequency components of the lung sounds. In this case, the JHUscope captured the

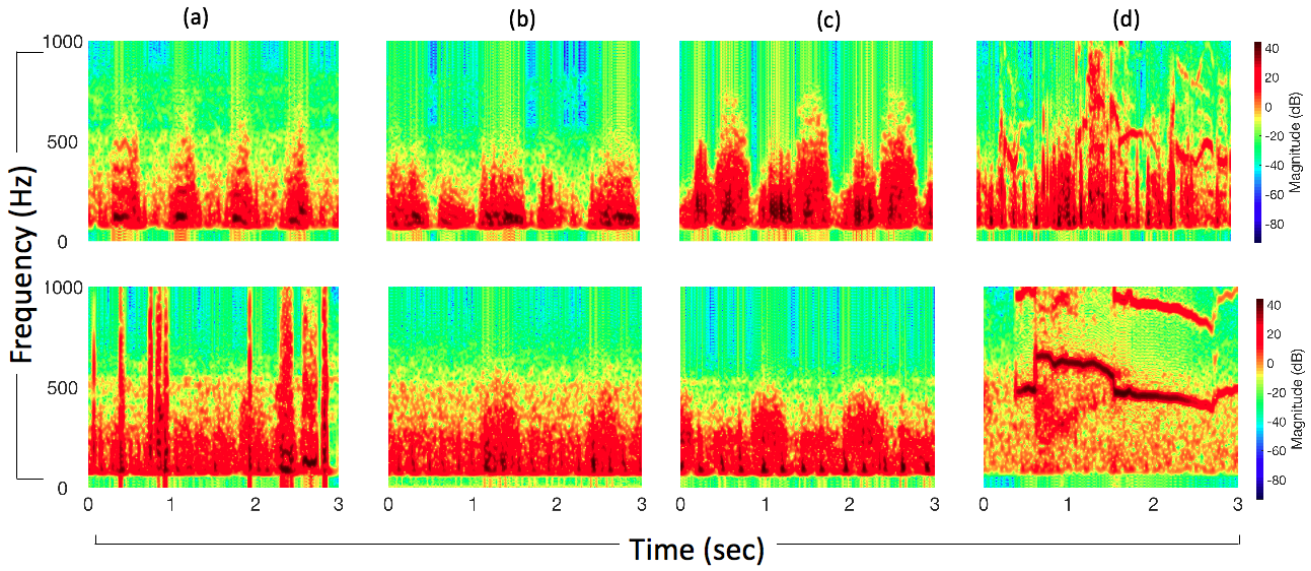

Fig. 5: Spectrograms of four patients with abnormal lung sounds. Top: JHUscope. Bottom: Thinklabs One

content between 100-300 Hz, yet the Thinklabs device only captured content around 100 Hz. The JHUscope audio signal, contained an audible rattling sound that was not audible in the Thinklabs audio signal. Pair 5d shows an example of a patient who was crying during the auscultation. In the JHUscope, some evidence of the crying is there (i.e., high frequency components around 500 Hz) yet the lung sounds were still captured as well. Only the crying is audible in the Thinklabs device and completely masks the breathing sound, making such auscultation using the Thinklabs device impossible. While these remain limited examples of clinical implications of changes in signal quality captured through a digital device, they reflect the potential clinical impact of sensor tuning and sensitivity to ambient noise.

IV. Conclusion

This validation study looked at objective features, which capture both the spectral and spectro-temporal properties of the data, for comparing a new technology (the JHUscope) to a commercially available system (Thinklabs One), in a variety of clinical settings. When clinical settings are controlled, as in the quiet hospital data, both systems yield very comparable performances and signal characteristics. This comparison agreement validates that the new device, though using different microphone layout and integration, does deliver a suitable auscultation signal. However, the comparison in highly unusual clinical settings that include remarkable levels of ambient noise and distortions suggests that tuning of stethoscope devices needs to balance use case and efficacy. Under-controlling for noise leakage (as is the case of the Thinklabs device) does result in masking of clinical markers of lung sounds which likely impacts possibility of diagnosis using auscultations.

Acknowledgements The authors would like to thank all collaborators who facilitated the data collection, as well as the patients and families enrolled in this study. We would also like to thank the funding sources: National Institutes of Health grants R01HL133043 and U01AG058532.

References

[1] Anonymous. Eko Digital Stethoscope, 2019.

[2] Anonymous. Thinklabs One Digital Stethoscope, 2019.

[3] Paul Boersma. ACCURATE SHORT-TERM ANALYSIS OF THE FUNDAMENTAL FREQUENCY AND THE HARMONICS-TO-NOISE RATIO OF A SAMPLED SOUND. *Proceedings of the Institute of Phonetic Sciences*, 17(Proceedings 17):97–110, 1993.

[4] T Chi, P Ru, and S A Shamma. Multiresolution spectrotemporal analysis of complex sounds. *Journal of the Acoustical Society of America*, 118(2):887–906, 2005.

[5] F Dalmasso, M M Guarene, R Spagnolo, G Benedetto, and G Righini. A computer system for timing and acoustical analysis of crackles: a study in cryptogenic fibrosing alveolitis. *Bulletin europeen de physiopathologie respiratoire*, 20(2):139–144, 1984.

[6] Eko Devices. CORE Digital Stethoscope - Electronic Stethoscopes, 2019.

[7] Mounya Elhilali and James E. West. The Stethoscope Gets Smart. *IEEE Spectrum*, 56(02):36–41, 2 2019.

[8] Laura E. Ellington, Dimitra Emmanouilidou, Mounya Elhilali, Robert H. Gilman, James M. Tielsch, Miguel A. Chavez, Julio Marin-Concha, Dante Figueroa, James West, and William Checkley. Developing a Reference of Normal Lung Sounds in Healthy Peruvian Children. *Lung*, 192(5):765–773, 10 2014.

[9] Laura E Ellington, Robert H Gilman, James M Tielsch, Mark Steinhoff, Dante Figueroa, Shalim Rodriguez, Brian Caffo, Brian Tracey, Mounya Elhilali, James West, and William 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*, 2(1):e000506, 2 2012.

[10] Dimitra Emmanouilidou and Mounya Elhilali. Characterization of noise contaminations in lung sound recordings. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2551–2554. IEEE, 7 2013.

[11] B Flietstra, N Markuzon, A Vyshedskiy, and R Murphy. Automated Analysis of Crackles in Patients with Interstitial Pulmonary Fibrosis. *Pulmonary medicine*, (2):590506, 2011.

[12] Arati Gurung, Carolyn G Scrafford, James M Tielsch, Orin S Levine, and William Checkley. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respir Med*, 105(9):1396–1403, 9 2011.

[13] Suneeti Nathani Iyer and D Kimbrough Oller. Fundamental frequency development in typically developing infants and infants with severe-to-profound hearing loss. *Clinical linguistics phonetics*, 22(12):917–936, 2008.

[14] Dror Lederman. Estimation of Infants' Cry Fundamental Frequency using a Modified SIFT algorithm. *Time*, pages 703–709, 2010.

[15] Hans Pasterkamp, Steve S. Kraman, and George R. Wodicka. Respiratory sounds: Advances beyond the stethoscope, 1997.

[16] A R A Sovij arvi, J Vanderschoot, and J E Earis. Standardization of computerized respiratory sound analysis. *European Respiratory Review*, 10(77):585, 2000.

[17] Vijee Venkatraman. MIT Researchers Develop the World's First USB-Powered Mobile Stethoscope, 2015.

[18] Jacob Williamson-Rea. Saving Lives with a Smart Stethoscope. *Johns Hopkins Magazine*, 9 2019.

[19] X Yang, K Wang, and S A Shamma. Auditory representations of acoustic signals. *IEEE Trans. Inf. Theory*, 38(2):824–839, 1992.