An objective measure of signal quality for pediatric lung auscultations

Annapurna Kala[1], Amyna Husain[2], Eric D McCollum[3], 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]Johns Hopkins School of Medicine, Global Program of Pediatric Respiratory Sciences,
Eudowood Division of Pediatric Respiratory Sciences, Department of Pediatrics

Abstract— **A stethoscope is a ubiquitous tool used to 'listen' to sounds from the chest in order to assess lung and heart conditions. With advances in health technologies including digital devices and new wearable sensors, access to these sounds is becoming easier and abundant; yet proper measures of signal quality do not exist. In this work, we develop an objective quality metric of lung sounds based on low-level and high-level features in order to independently assess the integrity of the signal in presence of interference from ambient sounds and other distortions. The proposed metric outlines a mapping of auscultation signals onto rich low-level features extracted directly from the signal which capture spectral and temporal characteristics of the signal. Complementing these signal-derived attributes, we propose high-level learnt embedding features extracted from a generative auto-encoder trained to map auscultation signals onto a representative space that best captures the inherent statistics of lung sounds. Integrating both low-level (signal-derived) and high-level (embedding) features yields a robust correlation of 0.85 to infer the signal-to-noise ratio of recordings with varying quality levels. The method is validated on a large dataset of lung auscultation recorded in various clinical settings with controlled varying degrees of noise interference. The proposed metric is also validated against opinions of expert physicians in a blind listening test to further corroborate the efficacy of this method for quality assessment.**

I. INTRODUCTION

A stethoscope is considered the most basic tool for the detection of pulmonary diseases since the 1800s. However, it remains a limited tool despite numerous attempts at reinventing the technology, due to major shortcomings including the need for a highly trained physician or medical worker to properly position it and interpret the auscultation signal as well as masking effects by ambient noise particularly in unusual clinical settings such as rural and community clinics.

With advances in digital technologies, some of these obstacles are being overcome. Recording and storing the lung sounds digitally paved the way to the development of computer-aided analyses in the field of auscultation. Several studies were focused on detecting adventitious breathing patterns[1], [2], [3]. Proper profiling of these pathological indicators could eventually be used in diagnosing pulmonary diseases thereby potentially substituting trained personnel in the lack of medical expertise.

New deep learning approaches have opened a lot of possibilities in fields like computer vision and speech recognition exploiting the availability of large amounts of data [4], [5], [6]. Access to data can also promote use of artificial-intelligence tools to aid diagnostics, telemedicine and computer-aided healthcare. In the domain of digital auscultations, the issue of data access and curation remains a limiting factor. While there are numerous studies that analyze lung sounds in laboratory settings or controlled environments [7], [8], [9], study conditions limit their applicability to real-life clinical conditions. Specifically, lung sounds collected in busy clinical settings tend to vary highly depending on the surrounding conditions at the time of recording [10]. Additionally, the differences in devices and sensors themselves exacerbate variability in the data collected. Ultimately, there are no agreed-upon standards as to what constitutes ”good data” in the domain of digital auscultations.

This study develops an objective metric of the quality of a lung sound. It is crucial to note that the metric is not an indicator of the presence or absence of adventitious lung sounds lending to the diagnosis or classification of lung sounds. Instead, it aims to deliver an independent assessment of the integrity of the lung signal and whether it is still valuable as an auscultation signal or whether it has been masked by ambient sounds and distortions which would render it uninterpretable to the ears of a physician or to an automated classification system.

One of the challenges for developing such metrics is the properties of breathing patterns like wheezes and crackles. In addition to covering a large frequency span of 50 to 2500 Hz between the two, these abnormal lung sounds often masquerade as noise. Any objective metric obtained should be careful about not misinterpreting such cases as low-quality. In this work, we try to achieve such a metric by working with both normal and abnormal lung sounds regarded to be of high quality by medical experts.

II. DATA

A. Data Acquisition and Preprocessing

Lung signals collected by the Pneumonia Etiology Research for Child Health (PERCH) study group [11] are used for the analysis. This study was conducted by International Vaccine Access Center, Johns Hopkins Bloomberg School of Public Health between 2011-2014 at 9 sites in 7 countries (The Gambia, Mali, Kenya, Zambia, South Africa, Bangladesh and Thailand). A Thinklabs digital stethoscope was used for collecting lung sounds from 8 body positions

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

with approximately 10-20 seconds per chest position. The clinical settings where the data was collected posed a number of challenges. Lung sounds were often masked by ambient noises such as background chatter in the waiting room, musical toys, vehicle sirens, mobile or other electronic interference. In addition, the study focused on pediatric pneumonia with all patients between 1-59 months old, which further exacerbated signal quality due to the subject's intense crying.

All signals were collected at 44.1KHz, and as part of pre-processing, low-pass filtered with a fourth-order Butterworth filter at 4 kHz cutoff, downsampled to 8 kHz, and centered to zero mean and unit variance. Signals were further enhanced to deal with clipping distortions, mechanical or sensor artifacts, heart sound's interference, subject's intense crying and ambient noise [12].

B. Data curation for quality assessment

We extracted 250 hours of recorded lung sounds from the PERCH dataset that were annotated by a panel of 9 expert listeners (pediatricians or pediatric-experienced physicians). Only segments for which a majority of expert listeners agreed on the clinical diagnosis (as normal or abnormal) with high confidence were kept. This curated subset of the data was considered to be a 'High Quality' database of auscultation signals for which there was a clear medical agreement from expert physicians on the patient's condition. We refer to this high-quality dataset collected in a everyday clinical settings as Γ_{HQ}. It included data from around 900 pediatric patients and contained an equal number of normal cases (no acute lower respiratory infections) and abnormal cases (signals containing crackles and wheezing which reflect acute lower respiratory infections including pneumonia).

To *systematically* vary the quality of this clean dataset, we corrupted these auscultations signals with ambient noises at controlled signal-to-noise (SNR) levels. Background noises consisted of sounds obtained from the BBC sound effects database [13], and included 2 hours of chatter and crowd sounds which comprised of wide range of noises like children crying, background conversations, footsteps and electronic buzzing. These BBC sounds effects signals were chosen as they offer non-stationary ambient sounds that reflect changes that can be encountered in everyday environments including clinical settings.

The entire Γ_{HQ} dataset was divided into Γ_{HQ}^{Train} and Γ_{HQ}^{Test} in a 80-20 ratio such that both datasets have equal number of normal and abnormal lung sounds. Γ_{HQ}^{Train} dataset was used to learn the profile of high quality lung sounds in an unsupervised fashion. Γ_{HQ}^{Test} was added to the BBC ambient sounds with varying signal-to-noise ratios ranging between -10 dB and 40 dB to obtain Γ_{Noisy} on which the quality metric was estimated.

Our final goal was to learn a regression model which estimates a quality metric based on the extent of corruption. For this purpose we formed a dataset $\Gamma_{Regression}^{Train}$ comprising 80% of Γ_{Noisy} having signal to noise ratios -5 dB, 10 dB and 20 dB. And to get a sense of perfect score, we also included 80% of Γ_{HQ}^{Test} in it. We tested the performance of the regression model on $\Gamma_{Regression}^{Test}$ which included the other 20% of Γ_{HQ}^{Test} as well as 20% of Γ_{Noisy} across all the signal to noise ratios ranging from -10 to 40 dB.

III. Methods

In this paper, we propose an objective quality metric for lung sounds which accounts for masking from ambient noise but is robust to the presence of adventitious lung sounds which are pathological indicators of the signal rather than a sign of low quality. We considered a wide set of low-level and high-level features in order to profile a clean lung sound (including both normal and abnormal cases), as outlined next.

A. Quality Metric Features

In order to estimate a quality metric, the following features were extracted from auscultation signals in Γ_{Noisy} dataset:

1) Spectrotemporal Features: An acoustic analysis of each auscultation signal was performed as follows: The time signal was first mapped to a time-frequency spectrogram using an array of spectral filters and following the approach proposed in Chi et al. [14]. This spectrogram was then used to extract four spectral and temporal characteristics of the signal as mentioned in [15]:

- Average spectral energy ($E[S(f)]$): This feature is obtained by averaging the expectation of energy content in the adjacent frequency bins of an auditory spectrogram.
- Pitch ($\hat{F_o}$): This fundamental frequency was calculated by matching the spectral profile of each time slice to a best fit from a set of pitch templates and estimating a maximum likelihood method to fit a pitch frequency to selected template. [16].
- Rate Average Energy ($E[\hat{R}(f)]$): This feature represents the average of temporal energy variations along each frequency channel over a range of 2 to 32Hz.
- Scale Average Energy ($E[\hat{S}(f)]$): These modulations capture the average of energy spread in the spectrogram over a bank of log-spaced spectral filters ranging between 0.25 and 8 cycles/octave.

2) Unsupervised embedding features: A convolutional neural network autoencoder was trained in an unsupervised fashion on $\Gamma_{HQTrain}$ dataset to obtain profile of high quality lung sounds which were considered clinically highly interpretable. As this dataset has equal number of normal and abnormal lung sounds, adventitious breathing patterns get represented as part of the 'high-quality' lung sound templates learned by the network; and are not considered as indicators of poor quality.

A three layer CNN was used as an autoencoder, and trained on auditory spectrograms generated from two second audio segments from the training dataset. The network learns filters that get activated if driven by certain auditory cues, thereby producing 2-dimensional activation map. The first two layers act as an encoder with the first layer extracting patches and second layer performing a non-linear mapping onto a low dimensional feature space; the third layer decodes the features back to the original spectrogram [17].

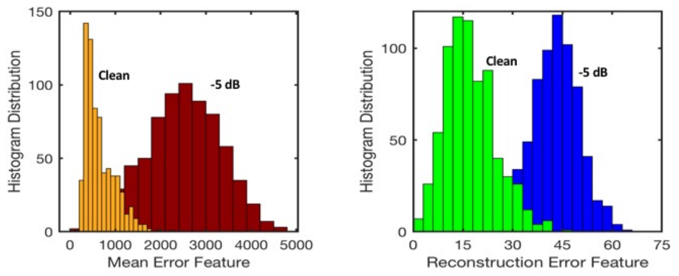

Fig. 1: Embedded features across different SNR values

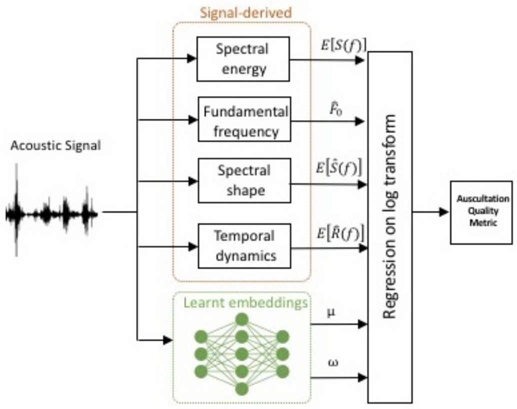

Fig. 2: Regression Block Diagram

Once trained, two parameters were extracted from this network, and used to supplement the signal-centered features in our measure of lung quality:

- Mean Feature Error (μ): After passing a spectrogram (32x128 dimensions) through the encoder (first two layers of the CNN), a dense low dimensional (32x32) embedding is obtained. An average of all the training CNN embeddings acted as a high-quality data low-dimensional 'template'. The L2 distance of the unsupervised features of the test data (Γ_{Noisy}) from the average feature template is taken as their corresponding Mean Feature Error. The left panel of Figure 1 shows the distribution of this mean error (μ) for high-quality signals in yellow. Overlaid on the same histogram is the distribution of mean errors obtained from -5 dB. The figure shows a clear shift in the mean feature error, indicating that it is a suitable attribute to supplement our quality metric.
- Reconstruction Error (ω): Assuming a good quality lung sound would be more similar to high-quality data and gives better reconstruction with the Autoencoder trained on clean data, we consider the L2 distance of the reconstructed spectrogram with the original spectrogram as the second embedding feature. The reconstruction errors of -5 dB SNR sounds exhibit a clear rightward shift from clean signals in the right panel of Figure 1.

B. *Quality metric*

Both signal-centric and learnt features (using the autoencoder) were combined together to yield an overall quality metric (Figure 2). The six features were integrated using a multivariate linear regression performed on the log transformation of the features. The regression labels for $\Gamma^{Train}_{Regression}$ ranged from 0 to 1 with 0 assigned to the -5 dB signal-to-noise ratio values and 1 to the un-corrupted lung sounds. 10 dB and 20 dB SNR audio clippings were given intermediate labels.

C. *Listening Experiment*

The quality metric obtained by regression was also validated by two expert physicians. The survey data included 92 lung sound recordings comprising clean data and signal to noise ratios of -5, 10, and 20 dB also corrupted with the BBC chatter and crowd noises. The subjects were asked to rate the quality of lung sounds on a scale of 1 to 5 with 1 being clinically completely uninterpretable and 5 being of the highest quality. All these lung sounds were shuffled in a random order ensuring a blind listening test.

As isolated two second segments used for the modeling were far too abrupt for human evaluation to assess the quality, we included entire audio clips with lengths varying from 10 to 20 seconds in the survey. The regression features for these longer signals were calculated by averaging all the features across their two second windowed segments.

IV. Results

The obtained quality metric shows a strong correlation of 0.8528 ± 0.0039 on a 10-fold cross validation across the span of signal to noise ratios with a high very high significance (p-value < 0.0001). The compliance of this correlation by lung sounds in $\Gamma^{Test}_{Regression}$ with additional signal to noise ratios which were not included in $\Gamma^{Train}_{Regression}$ further validates the quality metric as shown in Figure 3.

We also compared the quality metric scores with the expert evaluation of two physicians. We averaged the scores given by each of them as mentioned in Listening Experiment section. The quality metric also exhibits a high correlation of 0.7587 with the average expert score as observed in Figure 4.

V. Conclusions

Often times, we only have access to the recorded lung sound and not the surrounding ambient noise. This makes the estimation of noise content in the signal rather difficult. Since the lung sounds contain adventitious patterns which have similar spectral and temporal patterns as the ambient noise, it is difficult to gauge the quality by the signal alone. In this work, by creating a template of what a high quality lung signal sounds like irrespective of whether they are normal or abnormal(wheezes and crackles), we estimated a quality metric on a systematically corrupted database.

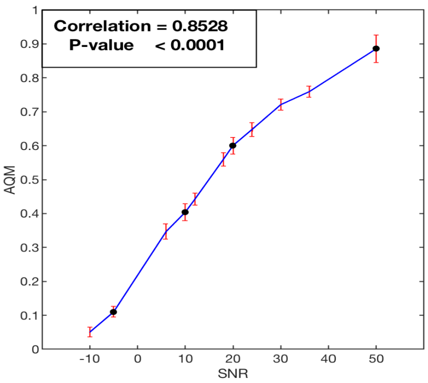

Fig. 3: Average Auscultation Quality Metric (AQM) from 0 to 1 vs Signal to Noise Ratio (SNR) in dB with the circles indicating the SNR values included in the $\Gamma^{Train}_{Regression}$. The error bars represent variance of AQM for each SNR.

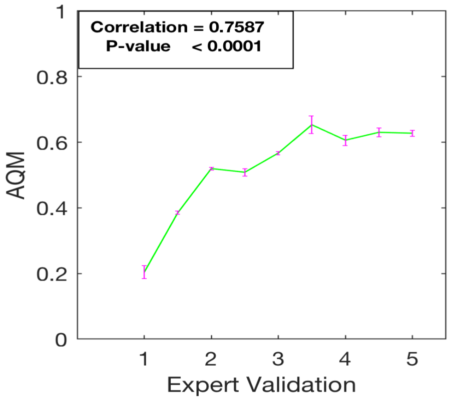

Fig. 4: Average Auscultation Quality Metric (AQM) from 0 to 1 vs Average Expert Score each scaled from 1 to 5 with variance of AQM for each score as error bars.

We used auditory salience features which account for the noise content as well unsupervised embedded features based on the clean template which justify the presence of the adventitious sound patterns. The obtained metric is also validated by the two expert physicians and the estimated quality score is on par with their evaluation. Further analysis could be done on testing the potential use of this metric as a preprocessing criteria for automated lung sound analyses. Also, if integrated with digital devices, data curation could be made more efficient by alerting the physician of the bad quality immediately to record again.

VI. ACKNOWLEDGEMENTS

The authors would like to thank the PERCH study group for guidance throughout the completion of this work, and to 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] M. Aykanat, Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *Eurasip Journal on Image and Video Processing*, vol. 2017, no. 1, 2017.

[2] D. Bardou, K. Zhang, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," *Artificial Intelligence in Medicine*, 2018.

[3] D. Chamberlain, D. Chamberlain, R. Kodgule, D. Ganelin, V. Miglani, and R. R. Fletcher, "Application of Semi-Supervised Deep Learning to Lung Sound Analysis Application of Semi-Supervised Deep Learning to Lung Sound Analysis," *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, no. August, pp. 804–807, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7590823/

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* IEEE, 2009, pp. 248–255.

[5] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 1, pp. 30–42, 2012.

[6] P. Pujol, S. Pol, C. Nadeu, A. Hagen, and H. Bourlard, "Comparison and Combination of Features in a Hybrid HMM/MLP and a HMM/GMM Speech Recognition System," *IEEE Trans.Speech and Audio Process.*, vol. 13, pp. 14–22, 2005.

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

[8] K. K. Guntupalli, P. M. Alapat, V. D. Bandi, and I. Kushnir, "Validation of Automatic Wheeze Detection in Patients with Obstructed Airways and in Healthy Subjects," *Journal of Asthma*, vol. 45, no. 10, pp. 903–907, 2008. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/02770900802386008

[9] 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.

[10] D. Emmanouilidou and M. 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).* IEEE, 7 2013, pp. 2551–2554. [Online]. Available: http://ieeexplore.ieee.org/document/6610060/

[11] E. D. McCollum, D. E. Park, N. L. Watson, W. C. Buck, C. Bunthi, A. Devendra, B. E. Ebruke, M. Elhilali, D. Emmanouilidou, A. J. Garcia-Prats, L. Githinji, L. Hossain, S. A. Madhi, D. P. Moore, J. Mulindwa, D. Olson, J. O. Awori, W. P. Vandepitte, C. Verwey, J. E. West, M. D. Knoll, K. L. O&#039;Brien, D. R. Feikin, and L. L. Hammit, "Listening panel agreement and characteristics of lung sounds digitally recorded from children aged 1–59 months enrolled in the Pneumonia Etiology Research for Child Health (PERCH) case–control study," *BMJ Open Respiratory Research*, vol. 4, no. 1, 6 2017. [Online]. Available: http://bmjopenrespres.bmj.com/content/4/1/e000193.abstract

[12] D. Emmanouilidou, E. D. McCollum, D. E. Park, and M. Elhilali, "Computerized Lung Sound Screening for Pediatric Auscultation in Noisy Field Environments," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 7, pp. 1564–1574, 2018. [Online]. Available: http://ieeexplore.ieee.org/document/7953509/

[13] BBC, "The BBC Sound Effects Library," 1990.

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

[15] N. Huang and M. Elhilali, "Auditory salience using natural soundscapes," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, p. 2163, 3 2017. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.4979055 http://www.ncbi.nlm.nih.gov/pubmed/28372080

[16] S. A. Shamma and D. J. Klein, "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," *Journal of the Acoustical Society of America*, vol. 107, no. 5, pp. 2631–2644, 2000.

[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2 2016.