ANALYSIS OF ACOUSTIC FEATURES FOR SPEECH SOUND BASED CLASSIFICATION OF ASTHMATIC AND HEALTHY SUBJECTS

Shivani Yadav[1]*, Merugu Keerthana*[2]*, Dipanjan Gope*[3]*, Uma Maheswari K.*[4]*, Prasanta Kumar Ghosh*[5]

[1]BioSystems Science and Engineering, Indian Institute of Science (IISc), Bangalore-560012, India
[2]Rajiv Gandhi University of Knowledge Technologies, Kadapa, Andhra Pradesh-516330, India
[3]Electrical Communication Engineering, Indian Institute of Science (IISc), Bangalore-560012, India
[4]Pulmonary Medicine, St. Johns National Academy of Health Sciences, Bangalore-560034, India
[5]Electrical Engineering, Indian Institute of Science (IISc), Bangalore-560012, India

ABSTRACT

Non-speech sounds (cough, wheeze) are typically known to perform better than speech sounds for asthmatic and healthy subject classification. In this work, we use sustained phonations of speech sounds, namely, /ɑː/, /iː/, /uː/, /eɪ/, /oʊ/, /s/, and /z/ from 47 asthmatic and 48 healthy controls. We consider INTERSPEECH 2013 Computational Paralinguistics Challenge baseline (ISCB) acoustic features for the classification task as they provide a rich set of characteristics of the speech sounds. Mel-frequency cepstral coefficients (MFCC) are used as the baseline features. The classification accuracy using ISCB improves over MFCC for all voiced speech sounds with the highest classification accuracy of 75.4% (18.28% better than baseline) for /oʊ/. The exhale achieves the highest classification accuracy of 77.8% (4.2% better than baseline). Comparable accuracies using speech sound /oʊ/ and non-speech exhale indicate the benefit of the rich acoustic features from ISCB. An analysis of 21 ISCB features groups using forward feature group selection shows that loudness and MFCC groups contribute the most in the case of /oʊ/, with interquartile range between 2^{nd} and 3^{rd} quartile of loudness feature being the best discriminator feature.

***Index Terms*—** Asthma, openSMILE, Classification, Sustained phonations,

1. INTRODUCTION

Asthma is a chronic airways disease which affects around 339 million people around the world with 1000 deaths every day [1]. Inflammation of airways causes chest tightness, breathlessness, wheezing, cough, and other unusual sounds during breathing in asthmatic subjects [2]. Diagnosis and monitoring of asthma is typically done by gold standard test Spirometry. Spirometry is a lung function test which measures how fast and how much air a person can exhale. During spirometry test, a patient has to take deep breath in and blow out air as fast as and as much as possible into the mouth piece, while nose is blocked with a nose clip. Values of Spirometry variables like forced expiratory volume in 1 sec (FEV1), forced expiratory capacity (FVC), and FEV1/FVC ratio are used to diagnose and determine the severity of asthma. Spirometry values depend on the efforts, coordination with technician and interest of the patients as it is very strenuous and tedious [3], especially for the children and old people.

Peak flow meter is another method which can be used for monitoring at home and ambulatory evaluation of asthma [4]. Peak flow meter measures peak expiratory flow rate (PEFR) from major airways but it fails to measure flow rate through minor airways which also get affected during asthma. Therefore, need for the time is an alternate technique which can overcome limitations of above mentioned methods. Sound based analysis can be one of the techniques which can help to alleviate the problem.

Sound based analysis is convenient for people irrespective of their age or medical conditions. This can be done by using speech sounds like sustained phonations and non-speech sounds like cough and wheeze. Cough is produced by contraction of expiratory muscles against a closed glottis and a sudden release of pressure afterwards [5]. Hiew et al. [6] analyzed acoustic characteristics of cough and used it for identification and counting of asthmatic coughs. Wheeze, on the other hand, is a whistling sound produced during breathing both inhale and exhale, due to obstruction in the airways. Study carried out by Wisniewski et al. [7] used spectral envelope and tonality index of wheeze in breath sounds for asthma monitoring. Lin et al. [8] performed wheeze detection based on neural network and spectrogram processing. In the literature, several works have used dominant frequency range [9], pitch [10], and duration of the breath [11] for asthma classification. Nabi et al. [12] used integrated power features of wheeze to detect asthma severity. A survey on wheeze detection of asthmatic patients can be found in the paper by Shaharum et al. [13]. Few of the reported works used time frequency spectrum [14], welch spectrum with feed forward neural networks [15], Gaussian mixture models with subband based cepstral [16], and power spectral density with autoregressive model [17].

Most of the works in the literature reported using wheeze and cough sounds for asthma detection and monitoring. Not much work in the literature exists that does speech based asthmatic and healthy classification. One of our research goals is to explore if asthma can be detected and monitored from spontaneous speech instead of cough, wheeze sounds. As people speak more often than having cough or wheeze, asthmatic signature, if present in speech, can be used to detect/monitor asthma in a passive manner.

In our previous work [18], we compared sustained phonations of five speech sounds, namely, /ɑː/, /iː/, /uː/, /eɪ/, /oʊ/, and non-speech sounds, namely, cough and wheeze for the asthmatic and healthy classification. Initial experiments are done with sustained phonations, instead of running speech as it modulates the voice due to co-articulation, making it more complex to understand the asthma signature in voice, if any. The work in [18] used six different statistics of used Mel-frequency cepstral coefficients (MFCC) with Support Vector Machine (SVM) as classifier We found that wheeze performs the best among all the sounds and /iː/ performs the best among sustained phonations.

In the present work, we expand the size of the database and set of speech sounds used for asthmatic patients and healthy subjects classification. We focus more on speech sounds by exploring acoustic features that could improve the classification accuracy. In particular, we have included unvoiced fricative /s/ and voiced /z/, as it has noise like structure, similar to the best performing stimulus wheeze.

Unlike our previous work using MFCC statistics, we used INTERSPEECH 2013 Computational Paralinguistics Challenge baseline (ISCB) [19] features comprising 21 feature groups for asthma and healthy classification. The goal of using ISCB features is to capture lot more variability in the speech sounds because these features consist of spectral, cepstral, energy, and excitation/source information [20]. It is known that signature of asthma is encoded in lung volume [21] and voice source is a function of lung capacity [22]. Dogan et al. [23] showed significant difference between the acoustic characteristics such as jitter, shimmer, harmonic to noise ratio of sustained phonations between asthmatic and healthy group. Similarly, Haman et al. [24] observed significant difference in the loudness of the running speech between asthmatic and healthy group. Both [24] and [23] observed significant Dysphonia between healthy and patient group, through assessment of speech language pathologist. Above mentioned discriminative information for both the classes can be captured by features present in the ISCB, which is not possible alone by MFCC statistics.

Experiments with 47 asthmatic and 48 healthy subjects demonstrate a significant improvement in all speech sound based classification accuracies by using ISCB features as compared to the baseline MFCC statistics features. We found /oʊ/ to perform the best with an accuracy of 75.4% using ISCB features compared to 57.12% using baseline features. However, non-speech sound exhale is found to be the best performing stimulus for classification using ISCB features with an accuracy of 77.8%, which is 4.12% (absolute) improvement over the baseline.

When forward feature group selection was performed, we observed that MFCC turns out to be the selected groups for most of the speech sounds. Apart from MFCC feature group, few other ISCB feature groups were also selected. For example, for sound /s/, spectral flux, spectral entropy, and loudness were found to be the best discriminative feature groups. A classification accuracy of 69.81% was obtained using /oʊ/ when only the best selected groups, namely, MFCC and loudness, were used. When the analysis of best performing groups was done further, we found interquartile range between 2^{nd} and 3rd^{rd} quartile of loudness feature to provide the maximum discrimination which results in a classification accuracy better than that using the baseline features. We begin with the description of the dataset.

2. DATASET

Dataset used in this experiment comprises 95 subjects with 47 (28M, 19F) patients and 48 (24M, 24F) healthy. Voice recording from patients and controls has been carried out at St. Johns National Academy of Health Sciences, Bangalore. Age of a healthy subject in this corpus varies over 19-60 years with an average age of 36 years. The patients are within the age group of 15-71 years with an average age of 43 years. Approval has been taken from the St. John's ethics committee for the recordings of the subjects. Consent form has been signed by the subject before the recording.

Subjects history and Spirometry test result have been used to group all subjects into healthy and patients under the supervision of doctors. Database consists of patients from different severity of asthma. Forced expiratory volume in 1 sec (FEV1) for 47 patients lies between 0.48 ls^{-1} to 3.59 ls^{-1} with an average of 1.53 ls^{-1} and standard deviation (SD) of 0.76 ls^{-1} . Similarly, FEV1/FVC for patients lies between 47%-97% of their reference values.

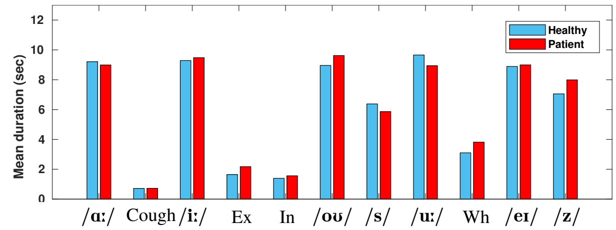

Fig. 1. Average duration of each stimulus. Wh, Ex and In denotes wheeze, exhale, and inhale respectively.

We recorded sustained phonation of speech sounds namely /ɑː/ (as in 'After'), /iː/ (as in 'Eat'), /uː/ (as in 'Cute'), /eɪ/ (as in 'Pay'), /oʊ/ (as Only'), /s/ (as in 'Same'), and /z/ (as in 'Zoom') and normal non-speech sounds cough and wheeze. Each stimulus was recorded on an average five times per subject. As wheeze could be present during either inhale or exhale, in addition to the entire breathing cycle.

The total number of recordings of speech sound are 1399 and 1583 for patients and healthy controls, respectively. Similarly, 1489 and 1576 non-speech sounds are there for patients and healthy subjects, respectively.

ZOOM H6 handy recorder was used for all recordings. The average time to complete recording of all stimuli is found to be 11.08 min. Data has been recorded in the spirometry lab of the hospital, which, in general, has a noisy background because of fan, AC noise and conversation between patients, technicians.

During recording, a nose clip is used to stop the air flow through the nose so that subjects can exhale through mouth to their full capacity. At the time of recording, sufficient breaks were given so that subjects were not tired due to the recording procedure. Start and end of each stimulus including inhale and exhale boundaries were marked manually by listening and visual inspection in Audacity [25].

Summary of average duration of each stimulus from all subjects is plotted in Fig. 1, which shows, on average, patients take longer time to empty their lungs due to obstruction compared to healthy subjects.

3. PROPOSED ANALYSIS OF SPEECH SOUNDS

3.1. ISCB features

ISCB has two features sets : 1) Set A: 59 low level descriptors (LLD) on which 54 functionals like arithmetic mean, centroid, moments etc are computed. Set A also contains delta features of those LLDs on which 46 functionals such as percentiles, quartiles, moments etc. are applied. 2) Set B: Contains 6 LLDs of source/excitation and their deltas with 39 functionals applied on each of them. There are five global temporal statistics defined for voiced/unvoiced segments of speech, applied on fundamental frequency LLD of Set B.
Both set A and B use some common functionals as well as set specific functionals. In addition, few functionals are only applied to LLDs delta. Functionals description can be found in [20].

Set A has 5900 features, while set B has 473 features. Therefore, a total 6373 features are present in the ISCB. To make analysis computationally viable, Set A was split into 15 groups and Set B into

Table 1. List of ISCB groups (number of features) .

	SetA	SetB
G R O U P S	Loudness (100), Modulated loudness(100), Root mean square (RMS) Energy(100), Zero crossing rate (ZCR) (100), RASTA auditory bands (2600), MFCC (1400), Band energy (200), Spectral Roll Off (400), Spectral flux (100), Spectral centroid (100), Spectral entropy (100), Spectral moments (300), Spectral slope (100), Harmonicity (100), Spectral Sharpness (100)	Fundamental frequency(F_0) (83), Probability of voicing (78), Jitter (78), Jitter of Jitter (JJ) (78) Shimmer (78), logarithmic harmonic to noise ratio (LHNR) (78)

6 groups. Group name with its features count in bracket are shown in Table 1.

3.2. Proposed analysis

Both the baseline features and ISCB features are used for asthma and healthy subject classification. In addition to comparing the classification performance using these two types of features, we also analyze the best performing feature groups by forward feature group selection as well as best performing feature within the best group. Classification, feature group selection as well as analysis of best performing features are carried out separately for each speech stimulus in order to analyze stimulus specific acoustic characteristics for asthma and healthy subject classification.

Fold wise forward feature group selection was performed in a manner similar to the forward feature selection algorithm [26] with classification accuracy as the selection criteria. A classifier was trained with each feature group and tested on the validation set. After training with individual groups, group with maximum validation accuracy was selected. A new classifier was trained with this best performing group jointly with each feature group one by one from the remaining set of feature groups. The trained classifier was applied on the validation set to compute the classification accuracy. The feature group pair, which achieved the highest classification accuracy was declared as the best performing feature group pair. This process is repeated till all the 21 feature groups are covered. From the sequence of best selected feature groups, the feature group with the highest accuracy on the validation set is declared as the best selected feature group. We repeat the same procedure for each fold. To choose the best selected feature groups across all folds we selected the groups which are common in at least 4 folds (also referred as common best performing groups (CBPG)).

To analyse the best features in best performing group combination, classifier was trained with individual features and tested on the test set. All features in the best selected feature groups are ranked using the classification accuracy on the test set.

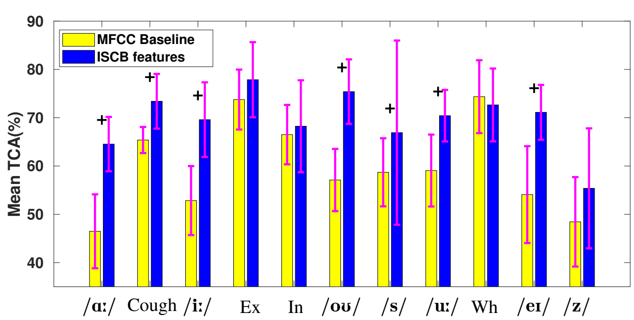

Fig. 2. Baseline and ISCB features comparison for all stimuli. Wh, Ex and In denotes wheeze, exhale and inhale, respectively.

4. EXPERIMENTAL SETUP

We used five-fold cross-validation setup for the classification experiments. Each fold was split into train, validation, and test sets. Number of train and test subjects for each fold is given in Table 2. Each fold contains around same number of patients and controls. Each train and test sets had same number of males and females from both the classes. 20% class balanced data from the train set was used as the validation set.

Table 2. Distribution of patients and healhty subjects in folds.

	Fold1		Fold2		Fold3		Fold4		Fold5	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Healthy	39	9	39	9	38	10	38	10	38	10
Patient	38	9	38	9	37	10	37	10	38	9
Total	77	18	77	18	75	20	75	20	76	19

We calculated ISCB features by using openSMILE (OS) toolkit for each stimulus. MFCC baseline was computed as mentioned in [18], by using Voicebox toolkit in MATLAB [27]. Features were normalized in the range $[0, 1]$. Features were calculated on stimuli level, it means for each instance of a stimulus we got 6373 dimensions feature. We used frame size of 20ms with 10ms shift for all the features except F_0 which uses 60ms window with 10ms shift. All features were smoothed with a window size of 3 frames. SVM classifier from LIBSVM toolkit [28] used as classifier. Hyper parameters, C was optimized by using Grid search method. Grid search was performed for $\log_2(C)$ in the range -6 to 8. We used the total classification accuracy (TCA) as an evaluation metric [18] because all folds were nearly class balanced.

Table 3. Best selected feature groups for all stimuli.

Stimuli	**Best selected feature groups**
/ɑː/	MFCC
/iː/	Jitter of the Jitter, MFCC
/oʊ/	Loudness, MFCC
/s/	Loudness, Spectral Entropy, Spectral Flux
/uː/	MFCC
/eɪ/	MFCC
/z/	Logarithmic Harmonic to Noise ratio

Table 4. % Reduction in features number by fold specific group feature selection.

Stimuli	**Fold1**	**Fold2**	**Fold3**	**Fold4**	**Fold5**	**Mean(SD)of TCA by using all 21groups**	**Mean(SD) of TCA by using best selected groups in each fold**
/ɑː/	72.37	76.46	68.01	67.32	67.39	62.95(6.83)	62.61(2.53)
/iː/	64.87	16.92	70.53	76.46	16.22	68.91(7.41)	64.99(5.29)
/oʊ/	74.89	56.06	74.89	72.1	28.17	74.18(5.26)	69.24(1.65)
/s/	34.18	82.13	65.82	86.22	94.07	66.82(18.03)	64.76(17.44)
/uː/	60.16	21.28	76.46	74.89	76.81	70.08(6.55)	65.21(12)
/eɪ/	36.01	69.31	67.66	76.46	76.81	69(8.17)	65.24(12.49)
/z/	82.13	85	50.4	70.45	94.07	56.3(8.15)	61.54(12.57)

5. RESULTS AND DISCUSSIONS

5.1. Baseline versus ISCB features

Comparison of mean TCA of five folds for all sounds using baseline and ISCB features is shown in Fig. 2. We observed that all stimuli showed an increase in mean TCA except wheeze, which

Table 5. List of name (Mean TCA %) of the best selected features in best groups. Here F1: MFCC, F2: loudness in Mel-scale, F3: MFCC_delta, F4: LHNR, F5: LHNR_derivative, F6: loudness in mel-scale_delta, QUA: Quartile, UPLT: Up-Level-Time, PER: Percentile, RQM: Root quadratic mean, AM: Arithmetic Mean, PAM: Peak mean, PR: Peak range, LRc2: linear regression 2^{nd} coefficients, POSAM: Positive arithmetic mean, PMRel: Peak mean relative to arithmetic mean, SKW: skewness, LPgain: linear prediction gain. In [ ] coefficient number is shown.

	Stimuli						
	/ɑː/(Mean TCA %)	/iː/(Mean TCA %)	/oʊ/(Mean TCA %)	/s/(Mean TCA %)	/u/(Mean TCA %)	/eɪ/(Mean TCA %)	/z/(Mean TCA %)
Best Selected Features	F1[2]_IQR2-3(61)	F1[1]_QUA1(63)	F2_IQR2-3(70)	F2_LRc2(73)	F3[8]_LPgain(66)	F1[3]_PER1.0(64)	F5_SKW(57)
	F1[2]_STD(61)	F1[8]_QUA3(63)	F2_PR(69)	F2_IQR2-3(73)	F3[9]_QUA1(65)	F1[11]_PMRel(63)	F4_QUA3(57)
	F1[8]_QUA3(61)	F1[1]_RQM(63)	F6_MRSlope(69)	F2_IQR1-3(72)	F3[8]_POSAM(65)	F1[11]_flatness(63)	F4_QUA2(57)
	F1[8]_UPLT50(61)	F1[1]_AM(63)	F2_iqr1-3(68)	F2_STD(72)	F3[10]_LPgain(64)	F1[8]_PER99.0(62)	F4_UPLT90(57)
	F1[6]_PER99.0(6)	F1[1]_PAM(63)	F6_STDRSlope(68)	F2_PER99.0(71)	F3[9]_IQR1-2(64)	F1[11]_LRc2(62)	F4_flatness(57)

showed decline in performance from 74.35% to 72.65%. All significant improvements, are indicated by + symbol at the top. Significance was tested by using Wilcoxon signed-rank test [29]. We observed that /oʊ/ (18.28% better than baseline) and /ɑː/ (18.07% better than baseline) are top two speech stimuli, which showed the maximum improvement. All the vowels showed an improvement of at least 10%. On the other hand, fricatives /z/ and /s/ improved by 6.94% and 8.2%, respectively. Although the maximum mean TCA of 77.87% is obtained using exhale among all sounds, /oʊ/ yielded a TCA comparable to this maximum. We observed among sustained vowels and fricatives, vowels performed better for the classification task.

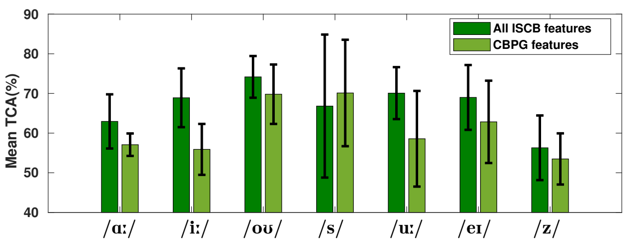

Fig. 3. Comparison of TCA by using all 21 groups of ISCB features and features from the best performing groups common across at least 4 folds (referred to as CBPG) for each stimuli.

5.2. Forward feature group selection

We observed a large drop in the number of features after fold specific feature groups selection. % features reduction is given in Table 4, which shows that the maximum % feature reduction occurred in Fold5 of /s/ and /z/. We observed that /z/ showed increase in mean TCA whereas other speech sounds showed a decline in performance. /ɑː/ showed nearly uniform reduction of 70% features in all folds and minimum drop of 0.34% in mean TCA.

The best performing group common across at least 4 folds are considered. They are referred to as common best performing groups (CBPG) is given in Table 3. From Table 3, we can see that MFCC was common in all vowels sounds. Mean TCA obtained using all ISCB features and features from CBPG are compared for each stimulus in the bar plot of Fig. 3 with standard deviation indicated by error bar.

We observed that all speech stimuli showed better performance with all 6373 features as compared to the CBPG except /s/. /iː/ sound showed the maximum reduction in TCA after best group selection, which consists of the jitter of the jitter and MFCC. On the other hand /z/ showed the minimum reduction in terms of TCA. It was observed that even the best group based TCA for each stimuli is higher than that using the baseline.

5.3. Analysis of best feature in best selected group

Top 5 features with maximum TCA on test set has been selected. Table. 5 shows the description of best selected features. We observed that MFCC 8^{th} coefficients statistics like inter quartile range between 2^{nd} and 3^{rd} quartile and 99^{th} percentile were common in /ɑː/, /iː/, and /eɪ/. For the best performing speech stimulus /oʊ/, out of five best performing features, three are from the Mel-scale loudness. We also observed that these best selected features were performing better than the baseline.

6. CONCLUSION

We observed that non speech stimulus is better than all the speech stimulus and /oʊ/ performed the best among all speech stimuli and second best among all speech and non-speech stimuli. All vowels showed a significant improvement in TCA with ISCB features as compared to the baseline features. Among fricatives, /s/ performed the best, while /z/ performed the worst among all the stimuli. With the CBPG features of speech sounds, we observed a drop in performance of all sounds as compared to using all ISCB features. MFCC is found to be the common best group across vowels sounds. Statistics of the MFCC 8^{th} coefficients carry more discriminative information. In future work, we want to explore the ISCB features computed from the running speech.

7. ACKNOWLEDGEMENT

Authors thank the Achuth Rao MV for his valuable inputs and the Department of Biotechnology, Govt. of India for their support in this work.

8. REFERENCES

[1] Global Asthma Network, "The global asthma report 2018. 2018," .

[2] FD Martinez, "Genes, environments, development and asthma: a reappraisal," *European Respiratory Journal*, vol. 29, no. 1, pp. 179–184, 2007.

[3] National Asthma Education, Prevention Program (National Heart, Lung, and Blood Institute). Second Expert Panel on the Management of Asthma, *Expert panel report 2: guidelines for the diagnosis and management of asthma*, DIANE Publishing, 1998.

[4] Bradley W DeVrieze and Steve S Bhimji, "Peak flow rate measurement," in *StatPearls [Internet]*. StatPearls Publishing, 2017.

[5] John N Evans and Marc J Jaeger, "Mechanical aspects of coughing," *Lung*, vol. 152, no. 4, pp. 253–257, 1975.

[6] YH Hiew, JA Smith, JE Earis, Barry MG Cheetham, and AA Woodcock, "DSP algorithm for cough identification and counting," in *Acoustics, Speech, and Signal Processing (ICASSP),IEEE*, 2002, vol. 4, pp. IV–3888.

[7] Marcin Wiśniewski and Tomasz P Zieliński, "Joint application of audio spectral envelope and tonality index in an e-asthma monitoring system," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 1009–1018, 2015.

[8] Bor-Shing Lin, Huey-Dong Wu, and Sao-Jie Chen, "Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network," *Journal of healthcare engineering*, vol. 6, no. 4, pp. 649–672, 2015.

[9] T Richard Fenton, Hans Pasterkamp, A Tal, and Victor Chernick, "Automated spectral characterization of wheezing in asthmatic children," *IEEE transactions on biomedical engineering*, , no. 1, pp. 50–55, 1985.

[10] Gérard Charbonneau, J Racineux, M Sudraud, and E Tuchais, "Digital processing techniques of breath sounds for objective assistance of asthma diagnosis," in *Acoustics, Speech, and Signal Processing, IEEE.*, 1982, vol. 7, pp. 736–738.

[11] Sandra Reichert, Raymond Gass, Christian Brandt, and Emmanuel Andrès, "Analysis of respiratory sounds: state of the art," *Clinical Medicine Insights. Circulatory, Respiratory and Pulmonary Medicine*, vol. 2, pp. 45, 2008.

[12] Fizza Ghulam Nabi, Kenneth Sundaraj, and Chee Kiang Lam, "Identification of asthma severity levels through wheeze sound characterization and classification using integrated power features," *Biomedical Signal Processing and Control*, vol. 52, pp. 302–311, 2019.

[13] Syamimi Mardiah Shaharum, Kenneth Sundaraj, and Rajkumar Palaniappan, "A survey on automated wheeze detection systems for asthmatic patients," *Bosnian journal of basic medical sciences*, vol. 12, no. 4, pp. 249, 2012.

[14] Kevin E Forkheim, David Scuse, and Hans Pasterkamp, "A comparison of neural network models for wheeze detection," in *IEEE WESCANEX 95. Communications, Power, and Computing. Conference Proceedings*, 1995, vol. 1, pp. 214–219.

[15] M Oud, "Lung function interpolation by means of neural-network-supported analysis of respiration sounds," *Medical engineering & physics*, vol. 25, no. 4, pp. 309–316, 2003.

[16] Mohammed Bahoura and Charles Pelletier, "Respiratory sounds classification using gaussian mixture models," in *Canadian Conference on Electrical and Computer Engineering*. IEEE, 2004, vol. 3, pp. 1309–1312.

[17] R Jané, D Salvatella, JA Fiz, and J Morera, "Spectral analysis of respiratory sounds to assess bronchodilator effect in asthmatic patients," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2000, vol. 6, pp. 3203–3206.

[18] Shivani Yadav, NK Kausthubha, Dipanjan Gope, Uma Maheswari Krishnaswamy, and Prasanta Kumar Ghosh, "Comparison of cough, wheeze and sustained phonations for automatic classification between healthy subjects and asthmatic patients," in *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 1400–1403.

[19] Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, et al., "The interspeech 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism," in *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association, Lyon, France*, 2013.

[20] Florian Eyben, *Real-time speech and music classification by large audio feature space extraction*, Springer, 2015.

[21] Ann J Woolcock, Anthony S Rebuck, John F Cade, and John Read, "Lung volume changes in asthma measured concurrently by two methods," *American Review of Respiratory Disease*, vol. 104, no. 5, pp. 703–709, 1971.

[22] Jenny Iwarsson, Monica Thomasson, and Johan Sundberg, "Effects of lung volume on the glottal voice source," *Journal of voice*, vol. 12, no. 4, pp. 424–433, 1998.

[23] Muzeyyen Dogan, Emel Eryuksel, Ismail Kocak, Turgay Celikel, and Mehmet Ali Sehitoglu, "Subjective and objective evaluation of voice quality in patients with asthma," *Journal of Voice*, vol. 21, no. 2, pp. 224–230, 2007.

[24] Abdul Latif Hamdan, Georges Ziade, Maher Kasti, Leslie Akl, Ibrahim Bawab, and Nadim Kanj, "Phonatory symptoms and acoustic findings in patients with asthma: a cross-sectional controlled study," *Indian Journal of Otolaryngology and Head & Neck Surgery*, vol. 69, no. 1, pp. 42–46, 2017.

[25] D Mazzoni and R Dannenberg, "Audacity [software]. pittsburg," 2000.

[26] A Marcano-Cedeno, J Quintanilla-Domínguez, MG Cortina-Januchs, and D Andina, "Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network," in *IECON 2010-36th annual conference on IEEE industrial electronics society*, 2010, pp. 2845–2850.

[27] Mike Brookes et al., "Voicebox: Speech processing toolbox for matlab," *Software, available [Mar. 2011] from www. ee. ic. ac. uk/hp/staff/dmb/voicebox/voicebox. html*, vol. 47, 1997.

[28] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.

[29] RF Woolson, "Wilcoxon signed-rank test," *Wiley encyclopedia of clinical trials*, pp. 1–3, 2007.