ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447657

AN UNSUPERVISED SEGMENTATION OF VOCAL BREATH SOUNDS

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

[1]BioSystems Science and Engineering, Indian Institute of Science (IISc), Bangalore-560012, India
[2]Electrical Communication Engineering, Indian Institute of Science (IISc), Bangalore-560012, India
[3]Pulmonary Medicine, St. Johns National Academy of Health Sciences, Bangalore-560034, India
[4]Electrical Engineering, Indian Institute of Science (IISc), Bangalore-560012, India

ABSTRACT

Breathing is essential to human survival, which carries information about a person's physiological and psychological state. Mostly breath sound boundaries are marked manually before being used for any task such as classification, spectral analysis, etc., which is very tedious. Various techniques have been proposed to segment breath sounds recorded at the chest, and trachea but vocal breath sounds (VBS) are under-explored. An unsupervised algorithm for VBS segmentation has been proposed in this work. Each breath phase in continuous breaths has been modeled using triangles, where the end points of triangles representing breath boundaries are estimated using dynamic programming. Data from 60 subjects (31 healthy, 29 asthmatic patients) having 307 breaths have been used. The proposed method's performance was found to be comparable with the manually marked boundaries. Comparable asthmatic versus healthy subject mean(standard deviation) classification accuracy using manually marked and predicted boundaries are 75%($\pm$ 11%) and 72%($\pm$15%), respectively are found.

Index Terms— Dynamic programming, Breath sound, Asthma, Segmentation, Vocal sounds

1. INTRODUCTION

Breathing is an irreplaceable process for human survival. Irregular breathing rate is one of the vital signs to indicate underlying poor psychological states like stress, anxiety as well as physiological conditions like cardiac arrest, asthma, COPD, etc. [1]. To determine lung health, breath sound analysis is an emerging technique among physicians and researchers. Breath sounds can be recorded from the chest (also referred to as lung sounds), trachea, nose, and mouth.

One of our research interest is vocal sounds(sounds which are recorded at the mouth) based asthma monitoring and diagnosis. First work in this thread was the classification between asthmatic and healthy subjects using sustained phonations, cough and breath sounds (also referred as VBS) [2], where among all the sounds, vocal breath sounds performed the best for the classification. Another work of ours shows that [3] the classification performance between asthmatic and healthy subjects is better with prior knowledge of breath boundaries as compared to randomly picked segments of breath signal from continuous breath cycles. In all our works breath sounds have been marked manually by the visual inspection of the spectrogram and listening, which is a very time-consuming task. In the literature, various methods such as by Palaniappan et al. [4], Aras et al. [5], Feng et al. [6], Yildirim et al. [7], Cam et al. [8] have been reported for the segmentation of breath sounds recorded at the chest and trachea but the segmentation of vocal breath sounds is least explored.

In this work, we propose an algorithm for the segmentation of breath sounds recorded at the mouth (referred as vocal breath sounds (VBS)). Unlike breath sounds at the chest and trachea, the microphone can easily record breathing sounds at the mouth with minimum effort and no physical contact with the patient. VBS data, we used in this work, is recorded in the hospital's natural noisy environment. The database consists of healthy and asthmatic patients' breath samples; hence, the proposed method also considers the variability of normal and abnormal breathing rates. As most of the breath signal information is present in a frequency range up to 2kHz, all breath samples are low pass filtered to 2kHz. The nearly periodic nature of the VBS energy has been exploited to find out the boundaries of VBS and its phases ('inhale' and 'exhale'). Each breath phase in continuous breaths has been modelled using triangles, where the end points of triangles representing breath boundaries are estimated using dynamic programming with prior breathing duration and number of breaths information. A method to estimate breath duration and the number of breaths has also been proposed in this work. An evaluation metric has been used to quantify, matched, missing, inserted, and deleted boundaries as given in the work by Ghosh et al. [9]. From the proposed method, we have found 89% boundary matched out of them 79% segment match with overlap rate [10] (mean(standard deviation)) of 88($\pm$13)%). Even the classification performance between asthmatic and healthy subjects using estimated boundaries found to be comparable with that of ground truth boundaries.

2. DATASET

For this work, data has been recorded from a total of 60 (24F, 36M) subjects, out of which 31 (12F, 19M) healthy controls and 29 (12F, 17M) patients. The subjects' age varies from 15 years to 53 years, where the average age of patients is 37.17 years, and for controls is 30.38 years. We have recorded data in the noisy condition of the hospital, which includes noises like people talking, fan, AC, phones ringing, etc. An informed consent form has been taken from each subject before recording. All patient recordings have been done in st. Johns Medical college hospital under the doctor's guidance. St. Johns medical college and hospital, Bangalore, Karnataka, India, Ethics Committee approved the study (Protocol number: IEC study ref no. 382/2018) on 12^{th}, February, 2019. All recordings have been done at the sampling rate of 44.1kHz and 16 bits using a ZOOM H6 handy recorder microphone. The microphone is kept at a distance of 3cm-5cm from the mouth while recording. While recording, the subject's nose is closed with the nose clip, breathing only through the mouth. Deep breaths of the subjects have been recorded. On average, 5 breath sounds have been recorded per subject. In total, 151

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on June 22,2024 at 08:42:46 UTC from IEEE Xplore. Restrictions apply.

controls' and 156 patients' breath sounds have been recorded. Hence the total 307. The average duration of breath sound is 2.97 secs, and the standard deviation is 1.47 secs. The minimum and maximum duration of breaths are 1.125 secs and 11.356 secs, respectively.

The two annotators have annotated recorded data by inspecting the spectrogram and waveform by using Audacity [11]. An example of an annotated waveform is given in Fig. 1. A noisy breath sample with 3 breaths is shown in Fig. 1. Inhale and exhale boundaries are marked separately (shown in red color) for each breath, and breath boundaries are shown in green.

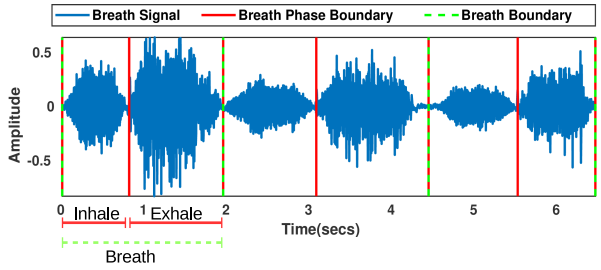

Fig. 1. Manually annotated breath sound sample file. Inhale and exhale boundaries are shown in green color and breath boundary is in red.

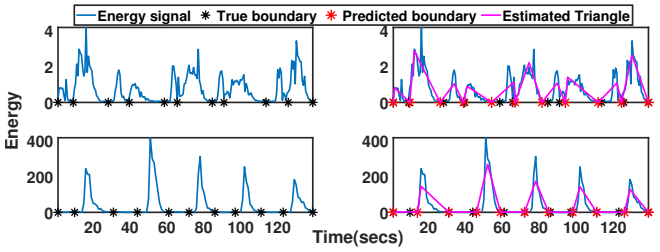

Fig. 2. Samples of energy signal. The left column shows the short-time energy signal with the ground truth boundary, and the right column shows the corresponding signal with the predicted boundary using triangle fitting.

3. PROPOSED BREATH SEGMENTATION

The Block diagram of our proposed method is represented in Fig. 3. The proposed method has three main steps. The first is the pre-processing step, which helps remove the high-frequency noise and estimates the signal's energy. The second step involves estimating average breath duration and number of breaths in a recording and using this information to find the boundaries using dynamic programming. Each part of the block diagram is explained below in detail.

3.1. Pre-processing

The energy of the breath signal (referred as $B[u]$, where u shows sample index), is present below 2kHz [12]. Therefore recorded breath sounds are low-pass filtered at cut-off 2kHz.Filtered signal has been framed with window size (w_s) and overlap (w_o). The energy of each frame has been calculated, and energy signal ($E[n]$) is obtained, where n denotes the frame index.

3.2. Breath Phase boundary prediction

3.2.1. Breath Phase Modeling

A breath has two components inhale and exhale. The energy of the breath signal taken from 4 subjects is shown in Fig. 2. In Fig. 2, left column shows energy signal with ground truth boundaries of breath phases marked, whereas the right column shows the corresponding predicted boundaries using the proposed method in addition to ground-truth boundaries. From Fig. 2 we can see that the energy signal is periodic in nature. However, the amplitude of short-time energy varies a lot due to the noisy nature of recordings and breathing patterns across subjects and within a signal itself. In this work, we utilize this periodic nature of the energy signal envelop shape. Each phase of breath is modelled by a triangle with varying duration where the endpoints of the triangle give the boundary. Hence, the short-time energy contour of a breath energy signal can be approximated by a sum of triangles, where number of triangles is equal to number of breath phases in the signal, and triangles, endpoints will give breath phase boundary. To understand how the triangle fitting has been done for each phase, consider a short-time energy contour $x[k]$ in a range of $k = 0, 1, 2, ...M$, where k denotes sample index. To fit a triangle to $x[k]$, between any three points $(k_1, 0)$, $(k_2, x[k_2])$ and $(k_3, 0)$, where $k_1 < k_2 < k_3$ is given by $F[k]$ where $x[k_2]$ is indicated as α. $F[k]$ is shown below.

$$F[k] = \begin{cases} \frac{\alpha(k-k_1)}{k_2-k_1}, & k_1 \leq k \leq k_2, \\ \frac{\alpha(k-k_3)}{k_2-k_3}, & k_2 \leq k \leq k_3 \end{cases}$$

To find the optimum (k_2, α) we need to minimize the following objective function.

$$J(k_1, k_3) = \underset{k_2,\alpha}{\text{minimize}} \sum_{k=k_1}^{k_2} (x[k] - \frac{\alpha(k-k_1)}{k_2-k_1})^2 + \sum_{k=k_2+1}^{k_3} (x[k] - \frac{\alpha(k-k_3)}{k_2-k_3})^2 \quad (1)$$

By differentiating Eq. 1 with respect to α and equating it to zero we get the following equation in terms of k_2 and α,

$$\alpha = \frac{\sum_{k=k_1}^{k_2} \frac{x[k]k}{k_2} + \sum_{k=k_2+1}^{k_3} \frac{x[k](k-k_3)}{k_2-k_3}}{\sum_{k=k_1}^{k_2} \frac{k^2}{k_2^2} + \sum_{k=k_2+1}^{k_3} \frac{(k-k_3)^2}{(k_2-k_3)^2}} \quad (2)$$

As from Eq. 2 we can see that (k_2, α) cannot be solved analytically, therefore k_2 varies from 1 to $k_3 - 1$ and at each given value of k_2, α has been calculated. Hence, for given (k_2, α) value of Eq. 1 can be calculated. From all calculated values of the objective function in Eq. 1, α and k_2 corresponding to the minimum value of will be the best fit.

3.2.2. Objective function for phase segmentation in breath signal

In this work, we have two assumptions. The first assumption is that breath starts from the first audio signal sample and ends at the last sample, which means there is no silence at the beginning and end of the signal. Secondly, the breath phases is continuous which means the endpoint of exhale is the beginning of the next inhale, and the endpoint of inhale is the beginning of next exhale. As explained in section 3.2.1, triangle fitting can be a good way to find the boundaries; therefore, we can say $E[n]$ is a train of triangles where consecutive triangles share boundaries. Hence finding the boundaries is

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on June 22,2024 at 08:42:46 UTC from IEEE Xplore. Restrictions apply.

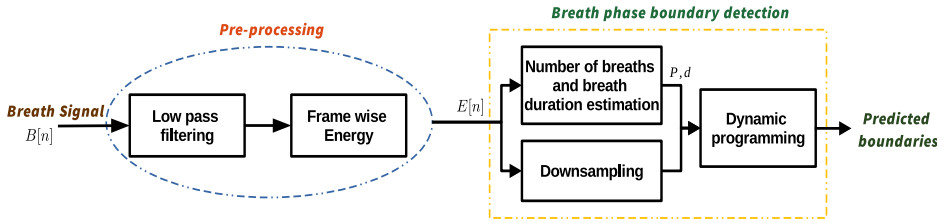

Fig. 3. Block diagram of our proposed method. $B[u]$, $E[n]$, P and d, indicates breath signal, energy signal, number of breath phases and breath phase duration, respectively. u and n, denotes samples index and frame index.

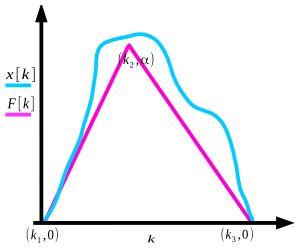

Fig. 4. Demonstration of triangle fitting.

formulated in terms of minimizing the following objective function. Let's assume if we know already number of phases (P) in the short-time energy signal $E[n]$ where $0 \leq n \leq M$, then breath phases boundaries, $n_1^*, n_2^*, ...n_{P+1}^*$ can be estimated by solving Eq. 3.

$$n_1^*, n_2^*, ...n_{P+1}^* = \underset{n_1, n_2 ... n_{p+1}}{\operatorname{argmin}} \sum_{p=1}^{P} J(n_p, n_{p+1}) \qquad (3)$$

$$\text{subject to} \quad n_1 = 0, n_{P+1} = M,$$
$$\{n_2, n_3 .. n_p\} \in \{2, ...M-1\}, n_{p+1} > n_p$$

Eq. 3 is the iterative cost function, which depends on the cost of fitting the previous breath phases; therefore, Eq. 3 can be solved using dynamic programming (DP) problem.

Steps to solve the Eq. 3 by using DP is given in Algorithm 1.

Algorithm 1 Breath phase boundary detection by solving equation Eq. 3.

Initialization:
P=number of phases, d=phase duration, δ=.3
$E_d[n]$=downsampled energy signal, $0 \leq n \leq M$
$O(1, n_2) = \min_{n_2} J(1, n_2), (Eq.1) \quad \lfloor d(1-\delta) \rfloor \leq n_2 \leq \lfloor d(1+\delta) \rfloor$
$I(1, n_2) = \arg\min_{n_2} J(1, n_2), \quad \lfloor d(1-\delta) \rfloor \leq n_2 \leq \lfloor d(1+\delta) \rfloor$
Recursion:

for k varies from 2 to P **do**
$O(k, n_{k+1}) = \min_{k\lfloor d(1-\delta) \rfloor \leq n_k \leq k\lfloor d(1+\delta) \rfloor} \{O(k-1, n_k) + J(n_k, n_{k+1})\}$
$I(k, n_{k+1}) = \arg\min_{k\lfloor d(1-\delta) \rfloor \leq n_k \leq k\lfloor d(1+\delta) \rfloor} \{O(k-1, n_k) + J(n_k, n_{k+1})\}$
$\forall\, n_{k+1} \in (k+1)\lfloor d(1-\delta) \rfloor \leq n_{k+1} \leq (k+1)\lfloor d(1+\delta) \rfloor$
end for
Back tracking:
$n_{P+1}^* = M$
for each phase (l) from P to 1 **do**
$n_l^* = I(l, n_{l+1}^*)$;

end for
return: $n_1^*, n_2^*, ..., n_{P+1}^*$

3.2.3. Number of breath phases (P)

As we explained in the previous block, boundaries of $B[u]$ can be found out by using DP, but DP requires information about the average duration of the breath phase (referred to as d) and Number of breath phases (referred as P). As $E[n]$ is a nearly periodic signal, a peak in the signal's magnitude spectrum should occur at this frequency of the signal. This property of $E[n]$ has been used to find P. From the spectrum of $E[n]$, the peak has been picked between our data's minimum and maximum breathing rate. The frequency at which peak has been picked is used as breath frequency from which P has been estitmated by doubling it. d is estimated by dividing length of $E[n]$ by P. These estimated values of d and P are used as input to DP.

Table 1. Inter-annotator agreement interms of Mean of Match(M), Insertion(I), Deletion(D), Segment match(S) and mean and standard deviation of Overlap rate(OvR) for segment matched breath sounds.

Total Boundaries	M(%)	D(%)	I(%)	S(%)	OvR mean(std)
367	97	3	3	82	92(11)

3.3. Reducing complexity

To reduce the time complexity of the DP, $E[n]$ is downsampled downsampled by a factor of 10 and the δ is used, which controls the search range around d. Both approaches reduce the time taken by the algorithm to find the boundaries to a great extent.

4. EXPERIMENTS AND RESULTS

To measure the performance of the algorithm, we calculated match, deletion, segment match, and insertion by comparing the locations of predicted and ground truth boundaries as described by Ghosh et al. [9]. If a ground truth boundary is in $\pm$ threshold of the predicted boundary, then it is called a match (M); otherwise, the ground truth boundary is considered deleted (D). Two consecutive matches are referred to as segment match (S). All the predicted boundaries which are not matched to any ground truth boundaries are known as inserted boundaries (I). All M, D, S, and I are given in percent.

Another evaluation metric, Overlap Rate (referred to as OvR) [10], quantifies how much segment-matched breaths and their predicted counterparts overlapped. OvR is defined as the ratio of the common duration between reference and predicted boundary divided by the maximum possible duration of the breath. OvR can lie between 1 (fully overlapped) and 0 (no overlap).

4.1. Inter-annotater difference

In the current work, breath, boundaries are annotated by two annotators. To find the agreement among annotators, one annotator marked breath boundaries is considered as reference and other annotator marked boundaries considered predicted. Mean of the M, I, D, S, and OvR has been reported. From Table 1, it can be observed that inter-annotator agreement is good. In this work, results are reported using only one annotator boundaries as ground truth.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on June 22,2024 at 08:42:46 UTC from IEEE Xplore. Restrictions apply.

4.2. Experimental settings

Breath sound signal ($B[u]$) has been low pass filtered at the cut-off of 2kHz by using the 6_{th} order Butterworth filter. $B[u]$ has been framed at $N_w = .1sec$ with $N_w = 10ms$ shift. The energy of each frame has been calculated to compute signal $E[n]$. To decrease the time complexity of DP, $E[n]$ is downsampled to 10 samples/sec from 100 samples/sec. $\delta = .3$ is set to reduce the complexity of DP. To estimate P and d, FFT of the $E[n]$ has been computed with FFT points twice the length of $E[n]$. The peak has been picked between the maximum and minimum breathing frequency of 0.833 Hz and 0.089 Hz from the $E[n]$ spectrum, respectively.

The classification setup is similar as given in [2]. Each train and test set have 50 and 10 subjects, respectively. Five out of six folds have 26 controls and 24 patients in the train set and 5 patients and 5 controls in the corresponding test set. The remaining one fold has 6 controls and 4 patients in the test set and 25 patients and 25 healthy subjects in the train set.

5. RESULTS

5.1. Comparison of proposed method between patients and healthy subjects

Table 2. Match(M), Insertion(I), Deletion(D), Segment match(S) and mean and standard deviation (in %) of Overlap rate(OvR) for segment matched breath sounds in patients and healthy.

Method used	Predicted Boundaries	M(%)	D(%)	I(%)	S(%)	OvR mean(std)
Patient	193	81	19	24	63	86(17)
Control	182	98	2	2	95	90(10)

As our data consists of both healthy and patients, we analyzed the proposed method performance between groups. Table 2 shows the results of the proposed method in healthy and patients. From the results, we can see that performance is better among the control group than patients as M is 98% and 81%, respectively. In the case of patients, S is 63%, which is very low compared to S of control, which is 95%. We also observe that in the case of patients, D and I are very high compared to healthy subjects. The reason for poor performance in the case of patients is due to incorrect prediction of P and hence d. Poor prediction of P can be due to irregular breath duration within continuous breath signal for a patient compared to healthy subjects due to breathing difficulty. However, for healthy subjects, the breath duration does not vary significantly.

5.2. Performance comparison with ground truth and predicted P and d

Results of breath segmentation by using predicted and ground truth boundaries P and d are shown in Table 3. From Table 3, it can be seen that M and S are higher by using ground truth boundaries as compared to estimated P, even though mean OvR is almost the same, being 88% and 90%, respectively.

Table 3. Match(M), Insertion(I), Deletion(D), Segment match(S) and mean and standard deviation (in %) of Overlap rate(OvR) for segment matched breath sounds using estimated and ground truth P and d.

P and d	M(%)	D(%)	I(%)	S(%)	OvR mean(std)
Estimated	89	11	13	79	88(13)
Ground truth	93	7	7	86	90(7)

The actual and predicted number of breaths in all 60 subjects are 307 and 315, respectively. Even though performance with ground truth P is high, it is not close to 100%. The reason behind that can be understood from the Fig. 5, where ground truth value of P lead to poor breath boundary prediction because of the first longer inhale sound than others. Hence, all breath boundaries got displaced. In this case, we got out of 7 breath boundaries 3 to be matched, 4 deleted, and 1 segment match. This example shows that our proposed method depends on accurate estimation of d and P.

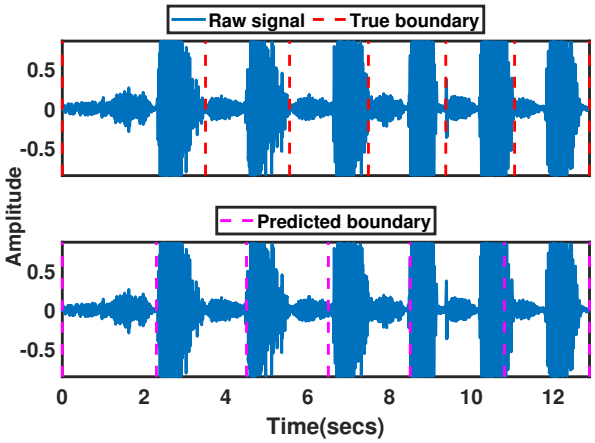

Fig. 5. Predicted boundaries using predicted P and d performed poor compared to ground truth P and d.

5.3. Asthmatic patients and healthy subjects classification

Fold-wise classification accuracy using estimated and ground-truth boundaries have been given in table 4. Mean TCA using estimated boundaries is 72%(± 15%), whereas, with ground truth boundaries, it is 75%(± 11%). Classification using estimated boundaries is close to the ground truth boundaries. Hence, the proposed method can be used for the segmentation, and segmented data can be used for the classification.

Table 4. Total classification accuracy (TCA)(%) between asthmatic patients and healthy subjects is shown for each fold using estimated boundaries and ground truth boundaries. Last column indicates the mean(std) of TCA averaged across all folds.

Breath Boundaries	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Mean (std)
Ground Truth	60	80	90	70	81.81	66.67	75 (±11)
Estimated	60	70	100	60	71.57	66.66	72 (± 15)

6. CONCLUSION

In this work, a vocal breath segmentation algorithm is proposed. The periodic nature of breath signal energy has been used to find the boundaries using dynamic programming. Predicted boundaries have good agreement with manually marked boundaries. Classification performance between asthmatic and healthy subjects is found to be comparable using estimated boundaries and ground truth boundaries. Future work includes robust estimation of breath phase duration and number of breath phases, estimation of breath phase boundaries having pause in between.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on June 22,2024 at 08:42:46 UTC from IEEE Xplore. Restrictions apply.

7. REFERENCES

[1] Thomas Ritz and Walton T Roth, "Behavioral interventions in asthma: Breathing training," *Behavior Modification*, vol. 27, no. 5, pp. 710–730, 2003.

[2] 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 Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 1400–1403.

[3] Shivani Yadav, Dipanjan Gope, K Uma Maheswari, and Prasanta Kumar Ghosh, "Role of breath phase and breath boundaries for the classification between asthmatic and healthy subjects," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021, pp. 870–873.

[4] Rajkumar Palaniappan, Kenneth Sundaraj, and Sebastian Sundaraj, "Adaptive neuro-fuzzy inference system for breath phase detection and breath cycle segmentation," *Computer Methods and Programs in Biomedicine*, vol. 145, pp. 67–72, 2017.

[5] Selim Aras, Mehmet ÖZTÜRK, and Ali Gangal, "Automatic detection of the respiratory cycle from recorded, single-channel sounds from lungs.," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 26, no. 1, 2018.

[6] Jin Feng, Farook Sattar, and Moe Pwint, "Application of walsh transform based method on tracheal breath sound signal segementation," in *2008 Proceedings of the First International Conference on Biomedical Electronics and Devices, BIOSIGNALS*, 2008, p. Volume 2.

[7] Isa Yildirim, Rashid Ansari, and Zahra Moussavi, "Automated respiratory phase and onset detection using only chest sound signal," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008, pp. 2578–2581.

[8] Steven Le Cam, Ch Collet, and Fabien Salzenstein, "Acoustical respiratory signal analysis and phase detection," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 3629–3632.

[9] Prasanta Kumar Ghosh, "Speech segmentation using extrema-based signal track length measure," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. IEEE, 2007, vol. 4, pp. IV–1065.

[10] Sérgio Paulo and Luís C Oliveira, "Automatic phonetic alignment and its confidence measures," in *International Conference on Natural Language Processing (in Spain)*. Springer, 2004, pp. 36–44.

[11] Dominic Mazzoni and R Dannenberg, "Audacity [software]," *The Audacity Team, Pittsburg, PA, USA*, vol. 328, 2000.

[12] Malay Sarkar, Irappa Madabhavi, Narasimhalu Niranjan, and Megha Dogra, "Auscultation of the respiratory system," *Annals of thoracic medicine*, vol. 10, no. 3, pp. 158, 2015.

Authorized licensed use limited to: J.R.D. Tata Memorial Library Indian Institute of Science Bengaluru. Downloaded on June 22,2024 at 08:42:46 UTC from IEEE Xplore. Restrictions apply.