



UNIVERSITY
of
GLASGOW

Qiu, Y. and Whittaker, A. and Lucas, M. and Anderson, K. (2005)
Automatic wheeze detection based on auditory modelling. *Proceedings of
the Institution of Mechanical Engineers, Part H: Journal of Engineering
in Medicine* 219(3):pp. 219-227.

<http://eprints.gla.ac.uk/2987/>

Automatic wheeze detection based on auditory modelling

Y Qiu¹, A R Whittaker^{1*}, M Lucas¹, K Anderson²

¹Department of Mechanical Engineering, University of Glasgow, Scotland, UK

²Department of Respiratory Medicine, Crosshouse Hospital, Kilmarnock, Scotland, UK

Abstract—Automatic wheeze detection has several potential benefits compared to reliance on human auscultation: it is experience-independent, an automated historical record can easily be kept and it allows quantification of wheeze severity. Previous attempts to detect wheezes automatically have had partial success, but have not been reliable enough to become widely accepted as a useful tool. In this paper an improved algorithm for automatic wheeze detection based on auditory modelling is developed, called the frequency and duration dependent threshold or fddt algorithm. Parameters of mean frequency and duration of each wheeze component are obtained automatically. The detected wheezes are marked on a spectrogram. In the new algorithm, the concept of a frequency and duration dependent threshold for wheeze detection is introduced. Another departure from previous work is that the threshold is based not on global power, but on power corresponding to a particular frequency range. The algorithm has been tested on 36 subjects, 11 of whom exhibited characteristics of wheeze. The results show a marked improvement in the accuracy of wheeze detection when compared with previous algorithms.

Keywords—breath sounds, wheezes, spectrogram, auditory modelling

1 Introduction

Auscultation by stethoscope is a quick non-invasive method widely used to assess pulmonary conditions. However, there are a few drawbacks. It lacks a method of recording, offers no quantitative description, and is experience-dependent. A computer-aided breath sound recording and analysis system has the potential to overcome these shortcomings.

* Corresponding author: Department of Mechanical Engineering, University of Glasgow, Glasgow, G12 8QQ, UK

Wheeze is a type of abnormal breath sound, which is a continuous adventitious sound having a musical character. Acoustically, wheezes are characterised by periodic waveforms with a dominant frequency (usually over 100Hz) and with a duration greater than around 100ms [1, 2]. In this study, wheeze is defined as an audible tonal signal in the frequency range 100-4000Hz of duration longer than 30ms. In this new definition, "audible" and "tonal" clarify the meaning of "musical character" used in earlier definitions. The lower frequency limit of 100Hz agrees with the literature definition, whereas the 4000Hz is limited by the sampling rate of 8000Hz. Since 10ms is the shortest span for tonal sound to be audible [3], the historical duration of 100 ms is too long to detect all audible wheezes. Though the limitation of the algorithm used in this work prevents the detection of signals with 10ms duration, it does allow durations of 30ms to be detected. This is therefore an improvement over earlier work. Wheezes are clinical signs of obstructive airway diseases if they are heard during spontaneous respiration or during induced airway narrowing. Previous research [4-7] has shown that, although the presence of wheeze is an insensitive sign for severity of airway obstruction, the occupation of wheeze in a breath cycle is a sensitive sign and therefore wheeze characteristics can be used to evaluate the severity of airway obstruction. Improving the accuracy of computerised detection will improve the ability to assess the severity.

It has been reported that, for the same set of recorded breath sounds, different specialists listening to the recordings may draw different conclusions and the same specialist may draw inconsistent conclusions under different conditions [8]. Computer analysis of breath sounds can remove this subjectivity from the data interpretation and a computerised system can be superior to the human ear, not only in its repeatability, but also in its ability to quantify the sounds in terms of frequencies and durations.

Previous automatic wheeze detection methods have shown that parameters relating to wheeze, such as frequency, duration and number of wheezes [4,6-11] and a description of how the wheeze evolves [11-13], can be extracted from breath sounds. As there is a lack of standard validation of wheeze detection methods, some of these published results were compared with a respiratory specialist's opinion of the wheezes present. From these comparisons, clinical examination, based on a combination of auscultation and visual examination of an expanded waveform of the breath sounds, was found to be superior to the wheeze detection algorithms in correctly identifying the wheeze content of wheezy breath sounds. Difficulties arise if the computerised system detects inaudible wheezes but, currently, there is no practical method of validating their presence.

Previous wheeze detection algorithms were based on a definition of a threshold above which wheeze signals, detected as peaks in the frequency domain, were distinguishable from a normal breath sound. These thresholds were specified in various ways. Fenton and associates' algorithm [5] used the criteria that a peak that was 15 times greater than the average power was a wheeze. Baughman and Loudon's algorithm [4,9] defined a wheeze as a peak with more than 3 times the baseline power. Homs-Cobera and co-workers [11,12] used a set of rules and a scoring system based on a set of empirical parameters, a development of Shabtai-Musih's work which proposed a less sophisticated set of rules [10]. What these studies have in common is a reliance on constant thresholds based on measures of global power to detect wheeze.

In this paper, a new wheeze detection method: the fddt (frequency and duration dependent threshold) algorithm is proposed. This differs from previous work in that the thresholds are defined based on auditory modelling. Instead of defining a constant threshold, the threshold in this new algorithm is frequency dependent. Also, rather than using global power, in this algorithm the threshold is based on energy of only part of a frequency range and is the energy within the selected auditory filter bandwidth. The choice of energy rather than power here is a result of previous work [3] which showed that the use of an energy threshold is more appropriate for detecting short duration sounds (less than 200ms).

2 Theoretical Basis of fddt Algorithm

2.1 Auditory modelling

The human's peripheral auditory system can be modelled as a bank of band-pass filters (auditory filters) with overlapping pass-bands [14]. A tonal sound is audible when its sound pressure level (SPL) is above the absolute hearing threshold and persists for more than 10ms [15]. When a listener is trying to detect a tonal sound that is masked by a noisy background, he is assumed to use a filter with a centre frequency close to that tonal signal. As the noise has a masking effect, it increases the SPL threshold at which the sound becomes audible, but only the frequency components in the noise within the auditory filter bandwidth have a masking effect.

The bandwidth of such an auditory filter is frequency dependent. A practical equivalent rectangular bandwidth (ERB) of such an auditory filter can be calculated as [14]

$$ERB_f = 24.7(4.37f + 1) \text{ Hz} \quad (1)$$

where f is the centre frequency in kHz.

2.2 Short-Time Fourier Transformation (STFT)

If the frequency content of a signal changes over time, then the standard Fourier transform cannot reflect the signal's time-varying nature. An easy way to overcome this issue is to compare the signal with elementary functions located in time and frequency simultaneously, that is [16],

$$STFT(t, \omega) = \int_{-\infty}^{\infty} s(\tau)h^*(\tau - t)e^{-j\omega\tau} d\tau \quad (2)$$

where s is the signal, h is a window function, h^* is the complex conjugate of the window function, t is shift in time, and ω is angular frequency. The window function, h , has short time duration. The Fourier transform of the signal s windowed with h shifted by t is calculated, thus the signal's local frequency characteristics can be characterised.

3 Procedures for Automatic Wheeze Detection

3.1 Sample data

All recordings of breath sounds were taken at the trachea. For convenience, mobile phones were used for remote monitoring of the subjects' breath sounds. Previous research has demonstrated that recordings by mobile phones were of similar quality to those recorded by microphones in the laboratory [ref lancet paper]. One data set was recorded from 20 subjects using the same mobile phone (Nokia model 5110) [17.]. Another data set was recorded from 16 subjects using the subjects' own mobile phones. To record a breath sound signal, the subject holds the mouthpiece of the mobile phone contacting the trachea and breathes normally for at least five full breath cycles. By dialling a number provided by a voicemail service, the acoustic signal captured by the mobile phone is sent to an e-mail address as an attached file. The attachment is a wave file in GSM 06.10 format (13 bit sampling at 8000Hz) with a time stamp. The wave files are then converted to 16 bit PCM format at 8000Hz. The mobile phone is not a recording device designed for tonal signals; adaptive signal processing is optimised for speech inputs. However, the non-speech performance appendix of GSM06.10 [18] states that sine waves in a frequency range of 100-2000Hz, with sufficient signal-to-noise ratios, will pass the encoding and decoding procedure. This is consistent with the results of recordings in this and the previous study [16].

3.2 The fddt Algorithm

Wheezes contain frequency components within the frequency range of normal breath sounds. Considering wheezes as additive tonal signals to a normal breath sound, and considering the normal breath sound as a masker, then the detection of wheezes is a task of detecting a signal in a masker. The tonal signal is therefore detectable when its energy is above the threshold of the masking signal.

The overall aim of the new fddt algorithm is to detect wheezes automatically. In this work an audible wheeze is identified as a peak in the frequency domain above the specified threshold.

The threshold is specified based on the experimental results of Reed & Bilger [3]. Their results have shown that the masking threshold is frequency and noise level dependant. To take account of the frequency dependency, the intensity of the signal energy must exceed that of the noise spectrum level (the average noise power per Hz) by amounts varying from 8dB at 250Hz to 14dB at 4kHz. The effect of noise level is less, making the above threshold values change by only around 1dB.

The baseline threshold level is given by

$$th = 9.5 + 3.48 \log_{10}(f) - 10 \log(ERB_f) \quad (3)$$

For a duration of around 100ms, if the ratio of signal energy to noise energy, E_S/E_N expressed in dB, is greater than or equal to th , then the signal is detectable. In this work the STFT resolution (Δt) is 32ms, so 96ms, or $3\Delta t$, is used as the baseline duration. Two more threshold levels are defined: $th + 3$, corresponding to 64ms, or $2\Delta t$, and $th + 4.8$, corresponding to 32ms, or Δt . Table 1 summarises this information.

Table 1 Threshold

| Duration (ms) | $10 \log E_S - 10 \log E_N$ (dB) |
|---------------|---|
| 96 | $th = 9.5 + 3.48 \log_{10}(f) - 10 \log(ERB_f)$ |
| 64 | $th + 3$ |
| 32 | $th + 4.8$ |

The procedure is illustrated in figure 1 and a more detailed explanation is given below.

1. Construct a time-frequency matrix A, which contains the power spectral density, by calculating the STFT with the following parameters: 512 samples, using a Hanning window of the same length with 50% overlap. Each cell of the matrix contains the power spectrum value ($|\text{STFT}|^2$) at a certain time interval $[t - \Delta t, t + \Delta t]$ and frequency band $[\omega - \Delta \omega, \omega + \Delta \omega]$. Initialise all elements of a binary matrix B to zero (the same size as the matrix A). The matrix B is used as a way of tagging the identified peaks.

2. Find peaks along columns of A, along the time axis. The peaks may either belong to a wheeze (signal) or noise. For the j th column of matrix A, if $A(i,j) > A(i+1,j)$ and $A(i,j) > A(i-1,j)$, then $A(i,j)$ is a peak.
3. Regarding a peak as corresponding to a signal component, calculate the energy of the signal, and the energy of the noise that passes through the auditory filter that centres on the signal. The filter bandwidth is calculated according to equation (1). If the signal-to-noise ratio is above the threshold, th , set the corresponding element of the B matrix to unity.
4. Label B using the connected component labelling (CCL) algorithm [19]. As a result, the same wheeze components have the same label. The CCL is used as follows:

Search the matrix B column by column and assign a value greater than 1 (a label) to each non-zero cell. The value is decided by the labels of the three neighbours (shown in figure 2). As a result of step 3, a maximum of only one of the three neighbours can be non-zero.

 - If all the neighbours of B (i,j) are 0, this means a new wheeze has been detected and $B(i,j)$ is assigned a new unused label.
 - If any one of the neighbours has a non-zero label, this means the cell $B(i,j)$ corresponds to the same wheeze as the previous non-zero neighbour, so the same label is assigned.
5. Referring to the length (duration) of signals with the same label, identify signals of less than 96ms duration and use the threshold levels of table 1 to remove the tags in the B matrix corresponding to shorter signals which do not exceed the higher threshold levels.
6. Breath sounds have a strong relationship with flow rate [20-23]. That is, all frequency components of spectral power are changed with flow rate: the higher the flow rate, the higher the spectral power. So, the wheezes and the noise can be considered as co-modulated by the flow. The co-modulation masking release phenomenon has the effect of decreasing the masking threshold [13]. This effect may mean that separately identified wheezes in step 5 are actually two components of the same wheeze. To compensate for this, detected peaks from step 2 which are neighbours of a wheezing component are relabelled so that they each have the same label, thereby connecting two wheeze signals. The neighbour definition is the same as in step 4.

7. Produce a spectrogram with wheezes marked. Calculate the average wheezing frequency (a wheeze usually has a frequency change from beginning to the end), standard deviation of wheezing frequency, duration of each wheeze, and percentage occupation of wheeze in each respiratory phase. Only average wheezing frequency and occupation in each phase are displayed to keep the results display compact, and these values can be saved in a database.

3.3 Implementing the fddt Algorithm

The algorithm is implemented in Matlab. A graphical user interface (GUI) has been developed to facilitate the following operations: load files, expand/restore parts of the graphs that are of interest, detect wheezes and play sounds. The GUI is shown in figure 3, and displays time domain plots, power spectrum plots and spectrograms corresponding to one cycle of a breath sound. Figures 3 and 4 show the GUI display of a normal and wheezy breath sound respectively. If wheezes are detected, they are marked with black pixels in the spectrogram, thus tracing the frequency evolution with time. A pop-up window shows the mean wheeze frequency and wheeze percentage occupation in each respiration phase as shown in figure 5. If the user wants to save the results, the results are exported to a Microsoft™ Access database file. If no wheezes are detected, the pop-up window shows 'no wheeze'.

4 Application of fddt Algorithm to Measured Breath Sounds

An example of the results of using the fddt algorithm is shown in figure 6, which displays contours of wheezes extracted from the breath sounds of a wheezy subject. The success of this algorithm is compared with three previously published algorithms for automatic wheeze detection. The three algorithms are applied to the same breath sound data as figure 6 and the results are shown in figures 7-9.

The length of sample data affects the frequency resolution, which can be seen from figures 6-9. The length of the minimum segments in the contours indicates the time resolution. Baughman's algorithm clearly has the lowest resolution, of 250ms. The window function can affect the

frequency accuracy too. According to Baraniuk and Jones [24], when the signal components resemble the window, excellent time-frequency representation of STFT can be achieved. A cosine window is used to calculate STFT in the fddt algorithm, which means that the frequency accuracy is good.

In the 20 subject data set, where recordings were from the same mobile phone, seven subjects had wheezes detectable by medical experts. Six patients had all wheeze components correctly identified by the fddt algorithm. A judgement that the wheeze components have been correctly identified is based on a combination of listening to the breath sounds, to estimate duration and pitch and to distinguish monophonic and polyphonic sounds, and interpreting the spectrograms to identify the visual wheeze lines in the time-frequency plane. Although this is not foolproof for real breath sounds, a similar procedure of comparison has been used for simulated breath sounds with known additive tonal signals representing wheezes, which has been successful in extracting the original wheeze signals using this algorithm [25]. This has leant confidence to the analysis of real subjects' breath sounds. The recordings from the seventh wheezy subject were of poor quality, and the fddt algorithm only detected some of the wheeze components. In the 16 subject data set, where subjects used their own mobile phones, four had wheezes which were correctly identified by the fddt algorithm.

The other three algorithms were applied to the same two data sets, for comparison. The resulting contours are therefore compared using different algorithms based on different criteria. However, the common criterion is the specialist's opinion, which has been a key component of the validation of all the published automatic wheeze detection algorithms. The other three algorithms were not as successful as the fddt algorithm. They managed to detect some of the wheeze components, but had more false detections of wheezes and failed to detect some true wheezes. In figures 7-9 the wheezes do not match the results in figure 6 and do not compare well with each other, mainly due to the specification of different constant thresholds in each of the algorithms. The higher thresholds missed the weak wheezes (as in figure 7), whereas the lower thresholds led to false detections (as in figure 8).

The comparison of results from the developed fddt algorithm with the results from others shows that the fddt algorithm is better. This is due to the benefits of the auditory model. The reliance on constant thresholds in the previously published algorithms is a fundamental problem. However, since the constant thresholds are data dependent, some improvements in wheeze detection from the data sets in this work may be achievable by making some adjustments to the constant thresholds used in these algorithms.

5 Conclusions

As wheezes (signals) are embedded in normal breath sounds (noise), for the automatic wheeze detection algorithms two key problems should be resolved: how to detect a wheeze and how to distinguish a wheeze. In the frequency domain, the answer to the first problem is to find a peak in the spectrum. The answer to the second problem is to define a threshold that ensures the peak is a signal peak.

The threshold definition in this work is different to previous threshold definitions in two aspects. One is that the threshold in this new algorithm is not a constant, but frequency dependent. This is reflected in the ERB_f calculation in table 1. Also, thresholds were previously based on global power in the calculated frequency range, while in this algorithm the threshold is based on energy of only part of frequency range, that is, the energy within the auditory filter bandwidth.

This algorithm provides an automatic wheeze detection tool, which has been shown to be more reliable than previous algorithms for data recorded from the trachea. The computer based system enables easy storage of historical data and results. The visual display facilitates reliable and easy detection of wheezes and the numerical presentation also facilitates quantification of wheeze severity, enabling easy management and comparison of historical records.

References

1. **Pasterkamp, H., Kraman, S.S., and Wodicka, G.R.** Respiratory sounds. Advances beyond the stethoscope. *Am. J. Respir. Crit Care Med.*, 1997, **156**, 974-987.
2. **Sovijarvi, A. R. A., Dalmasso, F., Vanderschoot, J., Malmberg, L.P., and Righini, G.** Definition of terms for application of respiratory sounds. *Eur. Respir. Rev.*, 2000, **10**, 597-610.
3. **Reed, C. M. and Bilger, R.C.** A comparative study of S/N₀ and E/N₀. *J. Acoust. Soc. Am.*, 1973, **53**, 1039-1044.
4. **Baughman, R. P. and Loudon, R.G.** Lung sound analysis for continuous evaluation of airflow obstruction in asthma. *Chest*, 1985, **88**, 364-368.
5. **Beck, R., Dickson, U., Montgomery, M.D., and Mitchell, I.** Histamine challenge in young children using computerized lung sounds analysis. *Chest*, 1992, **102**, 759-763.
6. **Fenton, T. R., Pasterkamp, H., Tal, A., and Chernick, V.** Automated spectral characterization of wheezing in asthmatic children. *IEEE Trans. Biomed. Eng.*, 1985, **32**, 50-55.
7. **Pasterkamp, H., Tal, A., Leahy, F., Fenton, R., and Chernick, V.** The effect of anticholinergic treatment on postexertional wheezing in asthma studied by phonopneumography and spirometry. *Am. Rev. Respir. Dis.*, 1985, **132**, 16-21.
8. **Pasterkamp, H., Wiebicke, W., and Fenton, R.** Subjective assessment vs computer analysis of wheezing in asthma. *Chest*, 1987, **91/3**, 376-381.
9. **Baughman, R. P. and Loudon, R.G.** Stridor: differentiation from asthma or upper airway noise. *Am. Rev. Respir. Dis.*, 1989, **139**, 1407-1409.
10. **Shabtai-Musih, Y., Grotberg, J.B., and Gavriely, N.** Spectral content of forced expiratory wheezes during air, He, and SF₆ breathing in normal humans. *J. Appl. Physiol.*, 1992, **72**, 629-635.
11. **Fiz, J. A., Jane, R., Homs, A., Izquierdo, J., Garcia, M.A., and Morera, J.** Detection of wheezing during maximal forced exhalation in patients with obstructed airways. *Chest*, 2002, **122**, 186-191.
12. **Homs-Corbera, A., Jane, R., Fiz, J.A., and Morera, J.** Algorithm for time-frequency detection and analysis of wheezes. 2000, 2977-2980.

13. **Waris, M., Helisto, P., Haltsonen, S., Saarinen, A., and Sovijarvi, A.R.** A new method for automatic wheeze detection. *Technol. Health Care*, 1998, **6**, 33-40.
14. **Moore, B. C. J.** An introduction to the psychology of hearing (4th edition). 1997, Academic Press.
15. **Yost, W. A.** Fundamentals of hearing-an introduction (4th edition). 2000, Academic Press.
16. **Qian, S. and Chen, D.** Joint time-frequency analysis : methods and applications. 1996, Upper Saddle River, N.J. : PTR Prentice Hall.
17. **Anderson, K., Qiu, Y., Whittaker, A.R., and Lucas, M.** Breath sounds, asthma, and the mobile phone. *Lancet*, 2001, **358**, 1343-1344.
18. ETSI EN 300 961 V8.1.1 (GSM 06.10) Appendix A.1.3
19. **Seul, M., O'Gorman, L., and Sammon, M.J.** Practical algorithms for image analysis : description, examples, and code. 2000,
20. **Harper, V. P., Pasterkamp, H., Kiyokawa, H., and Wodicka, G.R.** Modeling and measurement of flow effects on tracheal sounds. *IEEE Trans. Biomed. Eng.*, 2003, **50**, 1-10.
21. **Mussell, M. J., Nakazono, Y., and Miyamoto, Y.** Effect of air flow and flow transducer on tracheal breath sounds. *Med. Biol. Eng Comput.*, 1990, **28**, 550-554.
22. **Pasterkamp, H., Schafer, J., and Wodicka, G.R.** Posture-dependent change of tracheal sounds at standardized flows in patients with obstructive sleep apnea. *Chest*, 1996, **110**, 1493-1498.
23. **Soufflet, G., Charbonneau, G., Polit, M., Attal, P., Denjean, A., Escourrou, P., and Gaultier, C.** Interaction between tracheal sound and flow rate: a comparison of some different flow evaluations from lung sounds. *IEEE Trans. Biomed. Eng.*, 1990, **37**, 384-391.
24. **Baraniuk, R. G. and Jones, D.L.** Signal-dependent time-frequency analysis using a radially Gaussian kernel. *Signal processing*, 1993, **32**, 263-284.
25. **Qiu, Y.,** Measurement and analysis of breath sounds, PhD Thesis, University of Glasgow 2003.

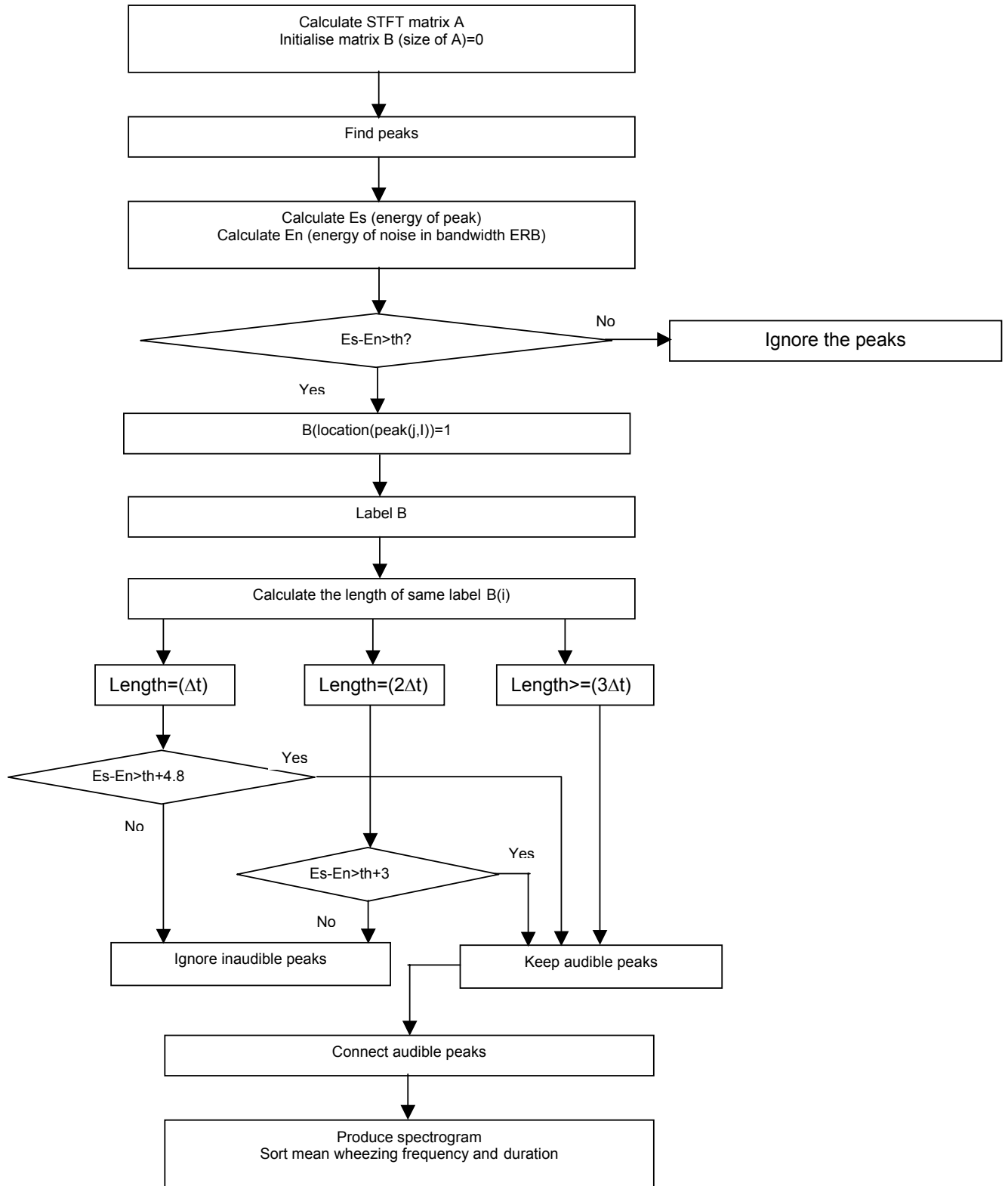


Figure 1 Schematic representation of fddt algorithm

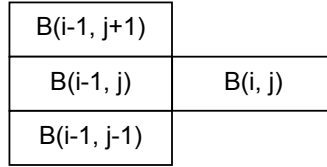


Figure 2 Neighbours definition in labelling algorithm

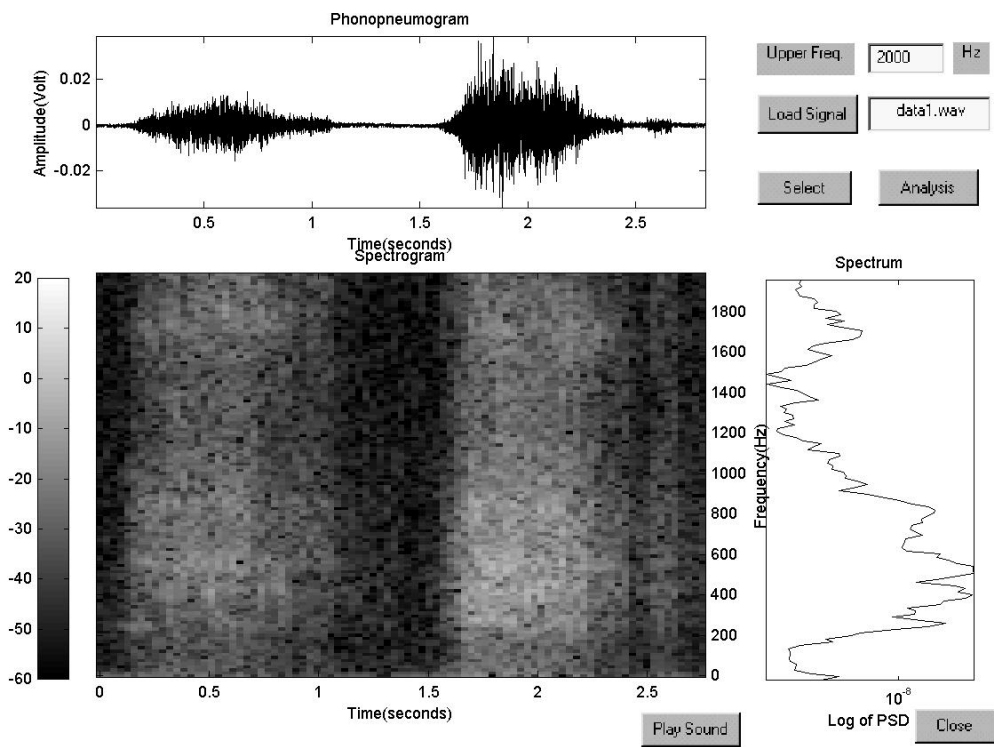


Figure 3 A normal tracheal breath sound of one cycle

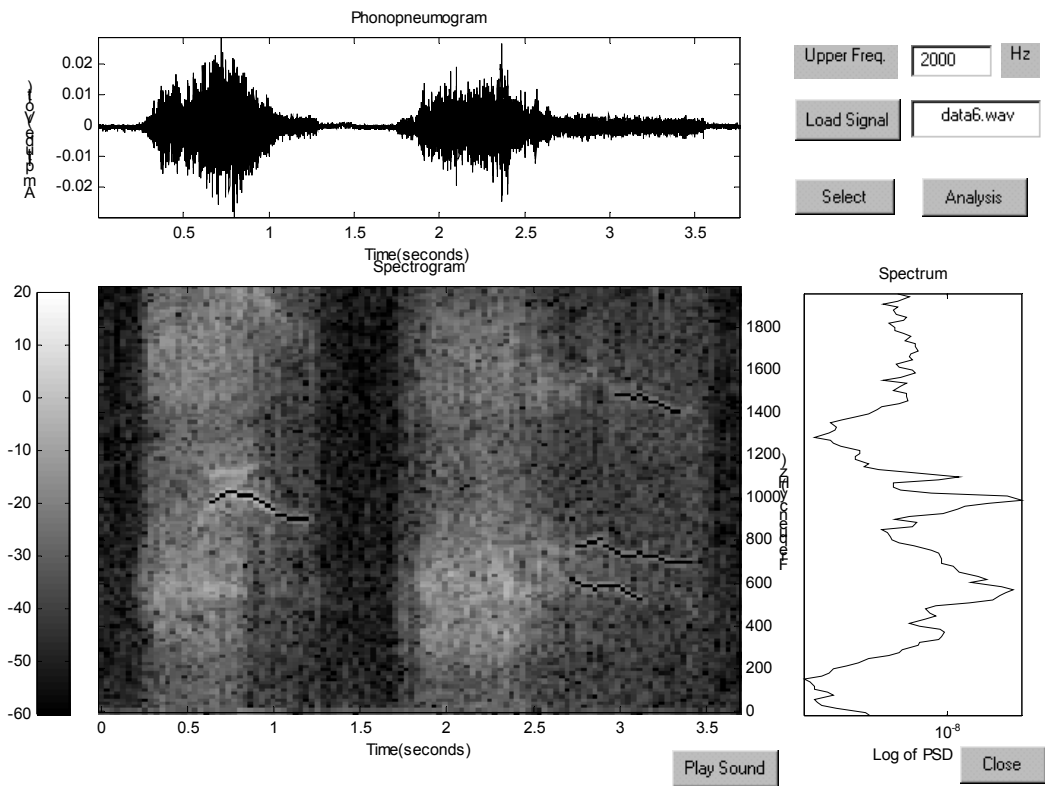


Figure 4 A wheezing tracheal breath sound of one cycle.

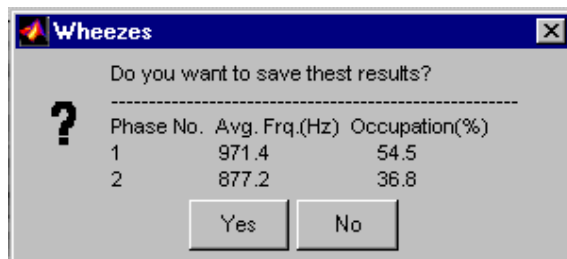


Figure 5 Pop-up window showing analysis results.

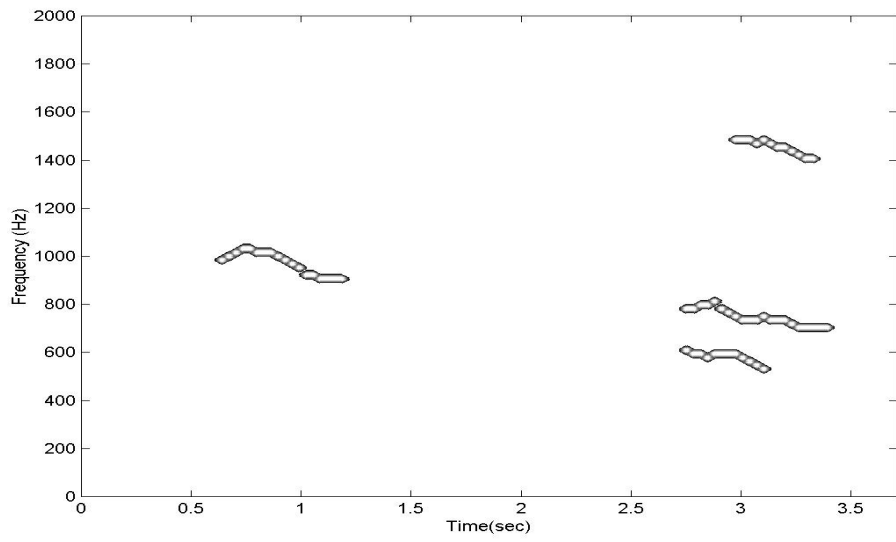


Figure 6 Contour of wheezes detected using fddt algorithm.

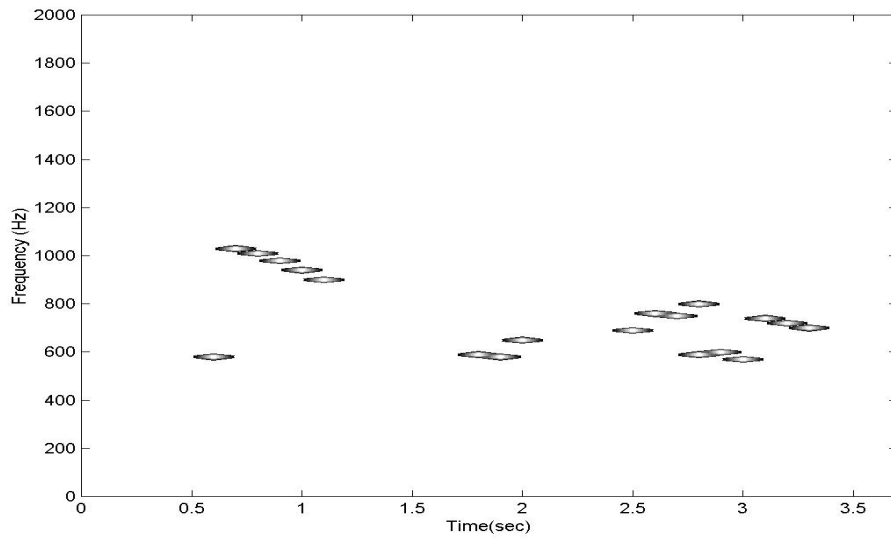


Figure 7 Contour of wheezes detected using Fenton's algorithm.

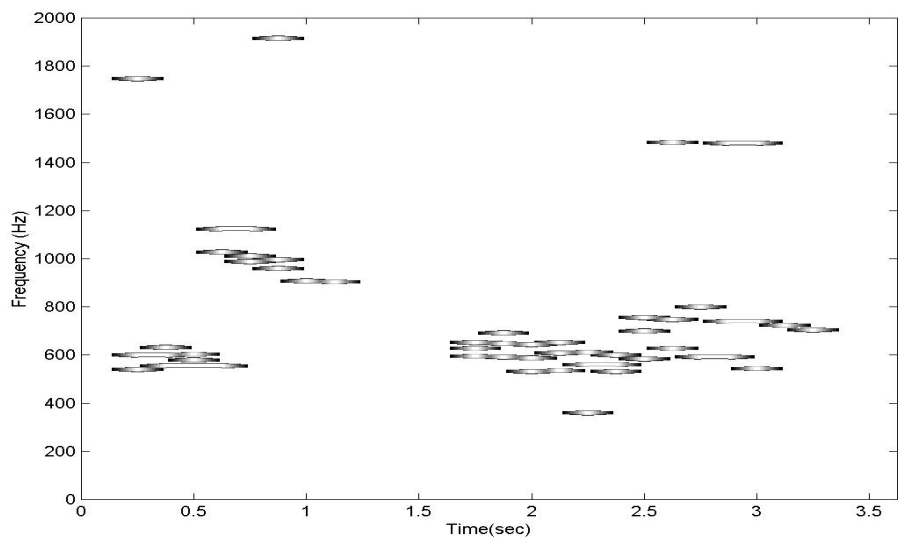


Figure 8 Contour of wheezes detected using Baughman's algorithm.

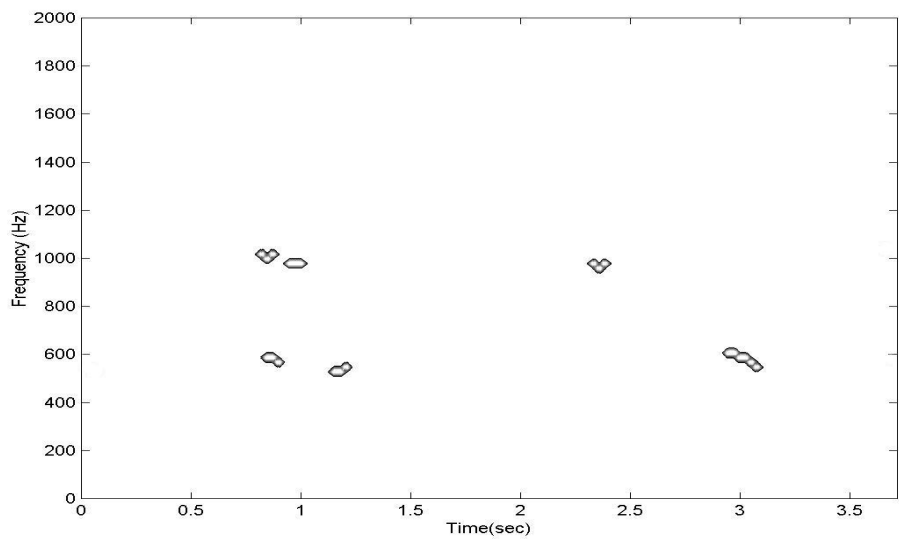


Figure 9 Contour of wheezes detected using Homs-Corbera's algorithm

List of figures

Figure 1 Schematic representation of fddt algorithm

Figure 2 Neighbours definition in labelling algorithm

Figure 3 A normal tracheal breath sound of one cycle

Figure 4 A wheezing tracheal breath sound of one cycle

Figure 5 Pop-up window shows analysis results

Figure 6 Contour of wheezes detected using fddt algorithm

Figure 7 Contour of wheezes detected using Fenton's algorithm

Figure 8 Contour of wheezes detected using Baughman's algorithm

Figure 9 Contour of wheezes detected using Homs-Corbera's algorithm

List of tables

Table 1 Threshold