Lung Sound Recognition Using Spectrogram and Adaptive Resonance Theory 2 Neural Network (ART2)

Achmad Rizal[1], **Vera Suryani**[2]
[1] *BioSPIN RG, IT Telkom Bandung,*
[2] *Informatics Department, IT Telkom Bandung,*
arl@ittelkom.ac.id ,vra@ittelkom.ac.id

Abstract

Usually, physicians diagnose lung diseases by listening to the lung sound using stethoscope. This technique is known as auscultation. Some lung diseases produce unique lung sounds, which refer to special recognized pattern. But the main problems concerning are the lung sounds that have low frequency (20 – 2000 Hz), low amplitude, in addition to other factors such as interference from other sounds, ear sensitiveness, and low variety of the pattern of lung sounds that make them almost similar. These came factors lead to the false diagnosing of lung disease if the auscultation procedures are not conducted correctly.

We proposed method to classify lungsound using spectrogram and ART2 neural network. By this method we can classify lungsound abnormality based on peak frequency each time. Experimental result shows accuracy of the system up to 98% for 70 testing data that consist of 5 classes of lung sound abnormality.

Keywords: Lung sound, auscultation, spectrogram, neural network, ART2

1. Introduction

Stethoscope as a diagnosis's tool tends to be subjectivity. The result depends on doctor ear's sensitivity and experience. Another disadvantage of this tool is that the voice could not be saved and could not be discussed with other doctors.

The electronic stethoscope is a solution. Lung sound can be heard, saved, and replay for further processing such as signal processing in order to reduce the noises for delivering clear voice. Additional technique is needed to recognize the voice as a result of auscultation process. This technique consists of lung sound acquisition processing, preprocessing, feature extraction, and classification. These digital signal processing can reduce the subjectivity of recognizing lung sound. But the software can not replace the doctor's function to deliver accurate diagnosis. This paper tries to build software to recognize the lung sound as a result of automatic auscultation process.

2. Theory

Respiratory system can be separated into two tracts ; upper and lower. The upper respiratory tract is comprised of nose, paranasal sinuses, pharynx, and larynx. The purpose of this tract is to purify, warm, and humidify ambient air before it reaches the gas exchange units. The lower respiratory tract begins with the trachea, the right main bronchus which divides into three lobes or divisions of the lung (upper, middle, and lower), the left main bronchus which divides into two lobes (upper and lower), followed by the bronchioles, and terminating at the alveoli (air sacs) which form the gas exchange surface [14].

Lung sound are generated by turbulent airflow through the respiratory tree.[10]. The turbulences happen when air flows from wider air cavity to narrower cair cavity and vice versa. Generally, lung sounds are classified into 3 groups: normal lung sound, abnormal and adventitious sounds. The sounds are separated into some catagories based on pitch, intensity, location and inspiration-expiration ratio[10]. Normal lung sounds are classified as tracheal, bronchial, bronchovesicular and vesikular sound.

Tracheal breath sounds are high-pitched and loud, with a harsh and hollow (or "tubular") quality. The inspiratory and expiratory phases are of equal duration, and there is a definite pause between phases. Tracheal breath sounds usually have very little clinical usefulness. Bronchial sounds normally are heard over the upper manubrium, these breath sounds directly reflect turbulent airflow in the main-stem bronchi. They are loud and high-pitched but not quite as harsh and hollow as tracheal breath sounds, the expiratory phase is generally longer than the inspiratory phase, and there is usually a pause between the phases. If bronchial sounds are heard over the thorax suggests lung consolidation and pulmonary disease. Pulmonary consolidation results in improved transmission of breath sounds originating in the trachea and primary bronchi that are then heard at increased intensity over the thorax.

Bronchovesicular sounds are normally heard in the anterior first and second intercostal spaces and posteriorly between the scapulas, where the main-stem bronchi lie. The inspiratory and expiratory phases are about equal in duration, with no pause between phases. Bronchovesicular sounds are soft and less harsh than bronchial breath sounds and have a higher pitch than vesicular sounds. Vesicular sound audible over peripheral lung fields, these breath sounds are soft and low-pitched, without the harsh, tubular quality of bronchial and tracheal breath sounds. The inspiratory phase is about three times longer than the expiratory, with no pause between phases.

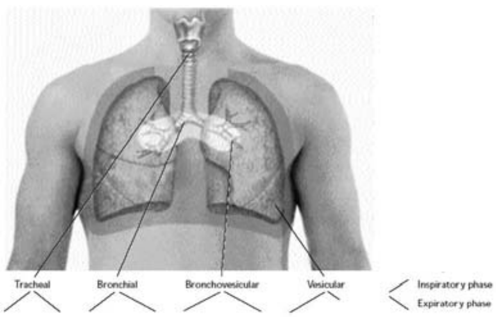

Figure 1. Location of normal lung sound [1]

In Addition, there are other lung sounds that produce by lung abnormality. These sound e.g : pleural rub, crackle, wheezing, grunting, and ronchi[8]. The sounds are classified by pitch, continuity, intensity. Usually lung condition must be analyzed by other inspection method such as palpation or percussion..

3. Material and method

The aim of this research is to design a lung sound recognition system that able to recognize lung sound automatically. Figure 2 shows the simplified block diagram of the system.

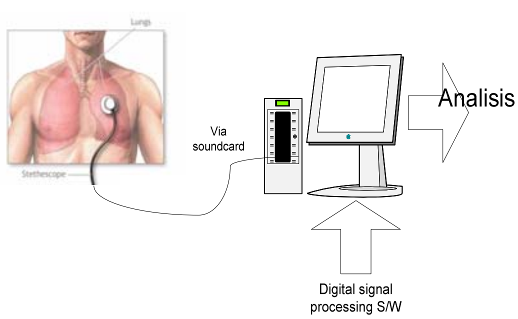

Figure 2. System design

3.1. Data

Input data for this system is recorded lung sound in *.wav, mono, sampling frequency at 8000 Hz. The data are collected from some resources in internet [2][8], then cutted into 1 respiration cycle. Data are consisted 5 class sown in table 1.

Table 1. Testing Data

Lung sound	
Bronchial	18
Medium crackle	15
Asthma	13
Friction rub	15
Velcro	9
Total	70

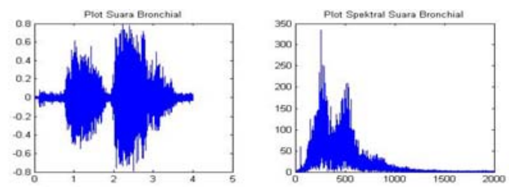

Figure 3. Bronchial lung sound

3.2 Signal Processing Methods

Generally, digital signal processing techniques can be shown in figure 4. The processes consist of preprocessing stage, feature extraction and classification.

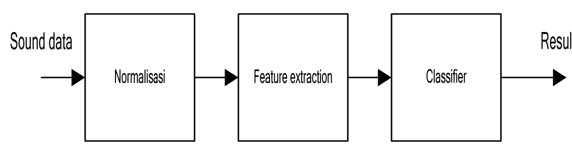

Figure 4. Signal processing process

3.3 Preprocessing

In order to make the signal has the same amplitude and remove the DC component, signal normalization is done. DC component removal is defined as follow.

$$S(i) = S(i) - \frac{1}{N}\sum_{i=1}^{N} S(i) \qquad (1)$$

where S(i) is the input signal in time domain. The next step is amplitude normalization. This process makes the maximum amplitude of signal become 1.

$$S(i) = \frac{S(i)}{S\ \max} \qquad (2)$$

3.4 Feature Extraction

We use spectrogram or Short Time Fourier Transform (STFT) to extract sound feature. Spectrogram is determined using the formula [5]:

$$STFT\ (t_n, f_k) = \sum_{l=t_n D-\frac{T}{2}}^{t_n D+\frac{T}{2}-1} w(t_n D - l)\, x(l)\, e^{-j2\pi f_k l/T} \qquad (3)$$

where x(l) is signal sample, w(tnD-l) is *time domain window* which its location is product of D samples. Spectrogram can be used to observe frequency component of signal every time. Resolution of spectrogram depends on window that use and overlap data between each window. In this research we use 512 points FFT and Keiser window with 500 sample length and 475 overlap samples. This process will produce 301 point STFT. Next process is calculating peak frequency of STFT that will produce 301 feature as input for ART2 NN.

3.5 ART2 neural network

ART2 (Adaptive Resonance Theory 2) is ART1 architecture improvement. These networks have differences in input data type at F1 layer. In ART1, input for layer F1 has binary value but ART2 input vector for F1 layer have continuous value[10]. ART 2 has modification method on input vector to accommodate continuous pattern. Because of the input vector has continuous value, that might has closed value, ART2 F1 layer architecture is more complicated than ART1. F1 layer in ART2 contains combination of normalization and noise reduction mechanism. The process need a reset mechanism to compare top-down signals and button-up signals. Figure 5 shows the architecture of ART2 neural network.

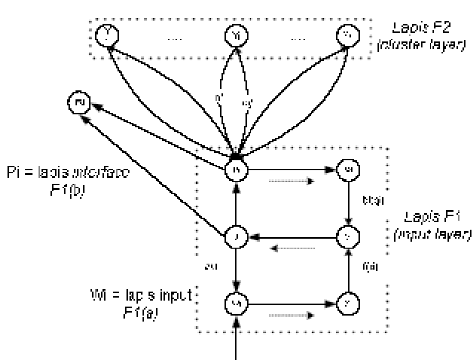

Figure 5. ART 2 Architecture [10]

F layer consist of 6 types of input unit : W,X,P,U,V and Q. Each unit consists of n unit that shows dimension of input pattern. Symbols and connection between units in F1 layer show the transformation that moving from one unit to another unit, it doesn't show multiplication for given value. Except for relation of Pi unit which is in F1 layer that shows multiplication weight of each transmitted signal between them.

F2 unit activation which become winner is d, where 1< d < 1. The activation function is applied on p vector and x vector concise with training algorithm. The symbol shows normalization process.

Competition occurs in F2 layer. Every unit Yi compete and winner-take-all, it means that the winner gets the learning process for each input unit from Pi in F1 layer. Learning process occurs only if the vector weight in up-down direction has sufficient similarity with the input vector. Activation function is given to input units from *Xi* (*f(xi)*) and unit *Qi* (*bf(qi)*). The activation unit acts to control some components of activation vectors which lay in lower stage which is chosen by user. Meanwhile in connection between W to U and from Q and V have fix value a and b.

4. Result and discussion

In our experiment, various parameters of feature extraction process and ART2 NN is observed to measure system performance.

4.1 Preprocessing result

Figure 6 and figure 7 show signal and frequency spectrum examples for bronchial and asthma, and normalization results. Figure 6 shows that bronchial sound has little DC component. It makes the normalization result just effect to amplitude of the signal. Figure 7 shows that asthma sound has significant DC component, so normalization process effect to signal amplitude and spectral.

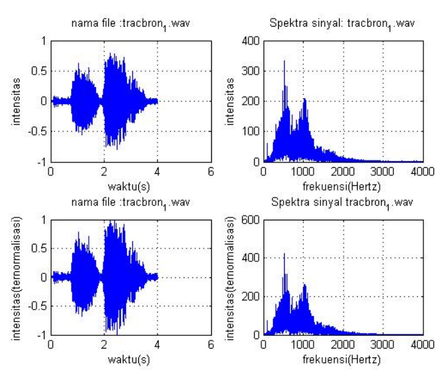

Figure 6. Bronchial sound, original and normalization result

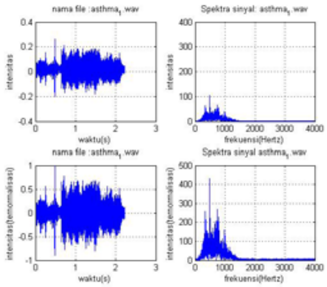

Figure 7. Asthma, original and normalization result

4.2 Feature Extraction

Actually, spectrogram of the signal must be calculated its each time peak frequency. This result will be input vector for ART2 NN. Spectrogram and peak frequency for bronchial sound are shown in figure 8-9

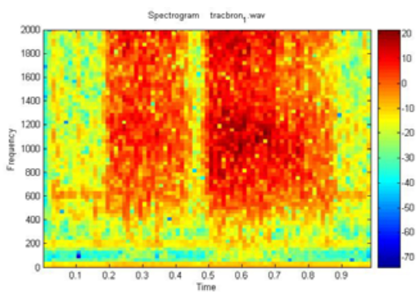

Figure 8. Spectrogram of bronchial sound

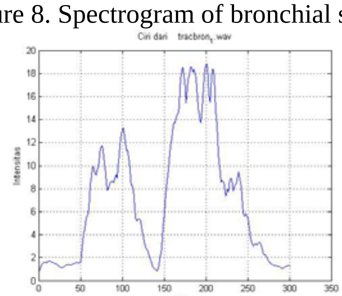

Figure 9. Bronchial sound’s features

4.3 ART2 classification result

Classification steps were conducted by changing ART2 parameter after each training session. ART2 parameter which produces the highest accuracy was chosen for next stage experiment. Accuracy is defined by :

$$\text{Accuracy} = \frac{\sum \text{correct recognition}}{\text{Total data}} \boldsymbol{x100\%} \qquad (4)$$

Correct recognition based on assumption active neuron definition. Each data will activate neuron and one class data shall activate one or more neuron. The most data class which is recognized in the neuron will be the neuron definition. So, another data class in the neuron will cause wrong recognition.

ART2 neural network was tested to recognized lung sound data with parameters which is shown in table 2.

Table 2. ART2 parameter

A =	10	θ =	0.02
B =	10	α =	0.1
C =	0.1	ρ =	0.997 - 0.999
D =	0.9	***iteration*** =	1-3
e =	0.0000001		

A, B, C, D, e, θ, and α, was set base on previous reserach [10] which is show high accuracy. Rho and iteration parameter were changed because these variables have significant effect to system performance, mainly accuracy. Higher rho value will make system more sensitif to input pattern. This condition effect to number of neuron that will be activated during training process. Taget of this reserch are high accuracy (>90%) but active neuron still in few number. Result for various rho values shown in table 3.

Table 3. Result for various rho values

rho	accuracy (%)	Active neuron
0.999	100	43
0.998	92.85	29
0.997	98.57	27

iteration = 3 , feature length = 301

For rho = 0.999, accuracy reach 100% but activated neuron reach 43 for 70 testing data. This mean each neurons recognize just 2 testing data, in other words system not too smart. For next observation, we choose rho = 0.997 because at this value, accuracy of the system quiet high and active neuron still fewer than other rho values. Result for various iteration values shown in table 4.

Table 4. Result for various iteration values

iteration	Accuracy (%)	Active Neuron
1	90	11

2	94.28	18
3	98.57	27

Rho = 0.997 feature length = 301

The best result reach when iteration = 3 but number of active neuron still high. For optimal result, we choose iteration = 2, because accuracy still high (94,28%) but has fewer active neuron. To observe effect of feature number to system performance, we conduct resampling process to the feature vector. We resample feature vector from 301 to 150 and 75, with rho = 0.997, iteration = 2. Result for this experiment shown by table 5.

Table 5. Result for various feature length

Feature	Accuracy (%)	Active Neuron
301	94.28	18
150	95.71	19
75	95.71	17

Rho = 0.997 , iteration= 2

Table 5 shows that accuracy still high when feature resampled until 75, and number of neuron does not changed significantly. For the optimum result, we can choose rho = 0.997, iteration = 2 and number of feature = 75.

4.4 Discussion

Spectrogram for feature extraction purposes need same data length. In lung sound case, this requirement will be difficult to be done. Because we can not record lungsound with the same length for different patient because it depend on respiration rate of each patient. For this purposes, we need 1 respiration cycles data fro input data. To solve this problem we can do resampling process, but this process will change information in the data. Other consideration to compute STFT of data is how to determine length of frame, number of overlap data for each frame, and determine type of window. This consideration will effect to resolution of STFT, and resolution will effect to feature that will be produced.

5. Conclusion

The experimental results demonstrate the proposed feature extraction methods (STFT and ART2 neural networks) have encouraging recognition performance. Accuracy of the system is about 95-98% for 5 class testing data. Further research is needed to improve performance of the lung sound recognition system.

6. References

[1] M. Kaelin, "Auscultation: Listening to Determine Dysfunction ". *Professionalization of Exercise Physiology online, An international electronic journal for exercise physiologists*. ISSN 1099-5862, Vol 4 No 8 August, 2001

[2] http://sprojects.mmi.mcgill.ca/mvs/mvsteth.htm

[3] T.S. Widodo "*Analisis Spektral Isyarat Suara Jantung*". Seminar On Electrical Engineering (SEE2004). hal 109-114, Universitas Achmad Dahlan, Yogyakarta, Agustus 2004

[4] L.C. Ludeman " *Fundamental of Digital Signal Processing*", John Wiley and Sons, 1987

[5] M.T. Pourazd, Z.K Mousavi, "*Heart Sound Cancellation from Lung Sound Recording Using Adaptive Threshold and 2D Interpolation in Time-Frequency Domain*", *Proc. IEEE Eng. Med. Biol Soci. (EMBS)*, pp 2586-89, Sept. 2003

[6] S. Salivahanan, A. Vallavaraj, C Gnanapriya., "*Digital Signal Processing*", McGraw-Hill, Singapore, 2001

[7] A. Rizal, "*Simulator Pengolahan Sinyal Biomedika*" Laporan Internship, Teknik Biomedika ITB, 2006

[8] www.rale.ca

[9] J. Saptaji, J Haryatno, A Rizal, "*Deteksi Kelainan Jantung Melalui Phonocardiogram (PCG) Menggunakan Metode Jaringan Saraf Tiruan Adaptive Resonance Theory 2*", Proceeding Tekno Insentif 2006, Kopwil IV, Bandung, Juli 2006

[10] A. Rizal, T R Mengko, A.B. Suksmono, "*Lung Sound Recognition Using Wavelet Packet Decomposition and ART2 (Adaptive Resonance Theory2) Neural Network*", BME Day 2006, International Conference on Biomedical Engineering, ITB Bandung, November 2006