Esophageal speech Recognition Utilizing Pulse Coupled Neural Network.

Fatchul Arifin(*), Tri Arief Sardjono(**), Mauridhy Hery Purnomo(**)
(*) jurusan Teknik Elektronika Univ Negeri Yogyakarta, fatchul@uny.ac.id
(**) jurusan Teknik Elektro Institut Teknologi Surabaya, t.a.sardjono@ee.its.ac.id, hery@ee.its.ac.id

Abstract

The laryngectomies patient has no ability to speak normally because their vocal chords have been removed. The simplest option for the patient to speech again is esophageal speech. Meanwhile, the voice recognition technology has been increased rapidly. In order the voice recognition technology also can be used by esophageal speech correctly, the esophageal speech recognition technology must be developed.

This paper describes a system for esophageal speech identification. Two main parts of the system, feature extraction and pattern recognition were used in this system. The Pulse Coupled Neural Network – PCNN is used to extract the feature and characteristic of esophageal speech. The pattern recognition, multi layer perceptron, will recognize the sound patterns.

From the experiments and results It can be concluded that the system can recognize esophageal speech very well up to 95,8 %. It is also can be known that PCNN can be utilized as feature extractor very well.

Keywords: *esophageal speech recognation, pulse coupled neural network (PCNN), multi layer perceptron (MLP)*

I. INTRODUCTION

The average number of laryngeal cancer patients in RSCM is 25 people per year [1]. More than 8900 persons in the United States are diagnosed with laryngeal cancer every year [2]. The exact cause of cancer of the larynx until now is unknown, but it is found some things that are closely related to the occurrence of laryngeal malignancy: cigarettes, alcohol, and radioactive rays.

Ostomy is a type of surgery needed to make a hole (stoma) on a particular part of body. Laryngectomy is an example of Ostomy. It is an operations performed on patients with cancer of the larynx (throat) which has reached an advanced stage. The impact of this operation will make the patients can no longer breathe with their nose, but through a stoma (a hole in the patient's neck) [3].

Human voice is produced by the combination of the lungs, the valve throat (epiglottis) with the vocal cords, and articulation caused by the existence of the oral cavity (mouth cavity) and the nasal cavity (nose cavity) [4]. Removal of the larynx will automatically remove the human voice. So post-surgery of the larynx, the patient can no longer speak as before.

Several ways to make laryngectomes can talk again have been developed. The easiest way is esophageal speech. Esophageal speech is a voice generated without the oscillation of the vocal folds. The voice is produced by releasing gases through the esophagus., in a manner similar to burping, to create speech. The esophagus functions in esophageal speech in much the same manner as the vocal cords in laryngeal speech, oscillating quickly to create distinct speech sounds. Esophageal speech is speaking by eructation [5].

Meanwhile research in the Speech recognition and its application is now going rapidly. A lot of application of speech recognition was introduced. Some of them are application of *voice recognition in cryptograph of public key* by magdalena [6], application of *voice recognition in* musical request by achmad basuki [7], application of *voice recognition in* car controller by ajub ajulian [8], and etc. Expected that this technology also can be used by esophageal speech satisfy.

This paper describes how to recognize the esophageal speech accurately by utilizing Pulse Code Coupled Network as speech feature extraction.

II. ESOPHAGEAL SPEECH SIGNAL PROCESSING

There are two main parts of the speech recognition systems; they are voice extraction and the pattern recognition. Voice extraction will take unique characteristic of the esophageal speech,

while pattern recognition is utilized to identify patterns voices.

In the year 2009, Muhammad Bahaoura compare the various methods related to feature extraction and pattern recognition for detection of diseases through the human respitory sound [9]. In the feature extraction he utilized some different methods, they are *Fast Fourier Transform* (FFT), *Linier Predictive Coding* (LPC), *Wavelet Transform* (WT), and *Mel-frequency cepstral coefficients* (MFCC). He also utilized some differents methods for pattern recognition proces, they are *Vector quantization* (VQ), *Gaussian Mixture Models* (GMM) and *Artificial Neural Netweork* (ANN). According to Bahoura, The combination between MFCC and GMM is the best methods related to the respitory sound.

Beyond what is presented by bahaoura, there is a feature extraction method that is used widely for image processing. This method is Pulse Coupled Neural network (PCNN).

PCNN is a binary model. Although it is very pupoler for image processing extraction, but now some researcher develop it for voice recognition. Taiji Sugiyama utilzed PCNN as pattern recognition unit. [10]

Esophageal speech recognition sistem which is proposed in this paper consist of *Fast Fourier Transform* (FFT), *Pulse Couple Neural Network* (PCNN), and *Multi Layer Perceptron* (MLP). Block diagram of this proces can be showed in Figure 1.

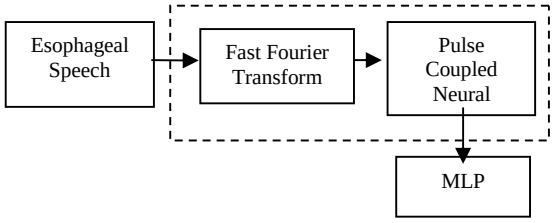

Fig. 1 Esophageal Speech Recognition process

Esophageal speech signal will be converted to the frequency domain by Fast Fourier Transform (FFT). This characteristic is important because the frequency domain gives a clearer view to be observed and manipulated than time domain.

Output of the Fast Fourier Transform will be sent to the pulse couple neural network for getting unique characteristic of esophageal voice.

The output of the PCNN will be fed into multi-layer perceptron (MLP). MLP will identify it, whether esophageal speech is recognized correctly or not.

III. PCNN FOR ESOPHAGEAL SPEECH RECOGNITION

PCNN is a pair of single layer neural network which is connected laterally, and two dimensions. Each of its inputs is connected to the input matrix. PCNN consist of:

- Input Part,
- Linking Part
- And Pulse Generator.

In the input part there are two parts, they are *linking input* and *feeding input*. Neurons receive input signal through the feeding and linking input [9]. Structure of Pulse Coupled Neural Network that utilized in this research can be seen in figure 2 below.

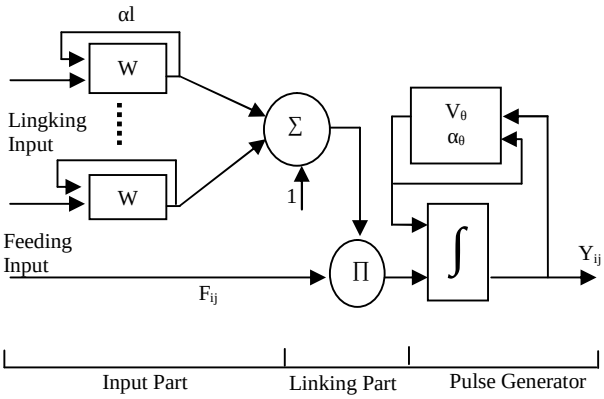

Fig. 2 Structure of PCNN

PCNN mathematical equations of the systems can be written as follows:

$$L_{ij}(n) = L_{ij}(n-1).\ e^{-\alpha L} + VL \quad (1)$$

$$U_{ij}(n) = F_{ij}(n).(1+\beta.L_{ij}(n)) \quad (2)$$

$$\Theta_{ij} = \Theta_{ij}(n-1).\ e^{-\alpha\Theta} + V_{\Theta} \quad (3)$$

$$Y_{ij}(n) = 2 \text{ if } U_{ij} > \Theta_{ij}(n) \quad (4)$$

$$1 \text{ if } U_{ij} = \Theta_{ij}(n)$$

$$0 \text{ otherwise}$$

Where:

Fij : feeding input,
Lij : linking input,
n : iteration number,
W dan M : Weight matrix,
* : Convolution operation,
Yij:output of neuron at coordinate i;j,
VL and VF : voltage potential,
α_L and α_F : decayed constants

Single signals of the linking input are biased and multiplied together. Input values Fij, and Lij are modulated in linking part of neuron. These proses will generate neuron internal activity Uij. If the internal activity is greater than dynamic threshold θij, the neuron will generate output pulses. In contrast, the output will be zero

The input matrix is transformed through PCNN into a sequence of temporary binary matrixes. Each of these binary matrixes has the same dimension as input matrix. The sum of all activities in specific iteration step gives one value which represents one feature for the classification.

It is evident that PCNN is not the neural network in the term of classification. It is only a mean of feature extraction for pattern classification.

IV.EXPERIMENTS AND RESULTS

In this paper, feature extraction of esophageal speech recognition has been conducted by PCNN. There are 16 ”A” vowel, and 8 ”B” consonant from different laryngectomes voice. Speech samples from data base were divided in to training and testing sets.

Recorded esophageal speech was sampled with sampling frequency 44.100 Hz and 16 bits resolution. It is assumed that the frequency of human voice signals is 300-3400 Hz. Sampling process must meet the Nyquist criterion. The nyquist criterion states:

$$f_s \geq 2xf_h$$

$$f_h = f_{in}Highest$$

Sampling frequency must be equal to or greater than twice input frequency. Thus, 44.100 Hz sampling has fulfilled the nyquist criterion.

Further, sampled signal will be converted from time domain in to the frequency domain utilizing *Fast Fourier Transform* (FFT). In this paper it is used FFT 512 point. Because the FFT is symmetric, the FFT output is taken only half of it. It is 256 data.

Output of FFT will be fed to the PCNN unit. In this paper PCNN used parameters as below:

$\alpha_L = 1$;
$\alpha_T = 0.2$;
$\beta = 3$;
$V_L=1.00$;
$V_T=20$;
L = zeros(p,q);
U = zeros(p,q);
Y = zeros(p,q);
Y0 = zeros(p,q);
Theta = zeros(p,q);

The some PCNN output of some ”A” vowel can be shown at figure 3. While PCNN Output of some ”B” consonant can be shown at figure 4.

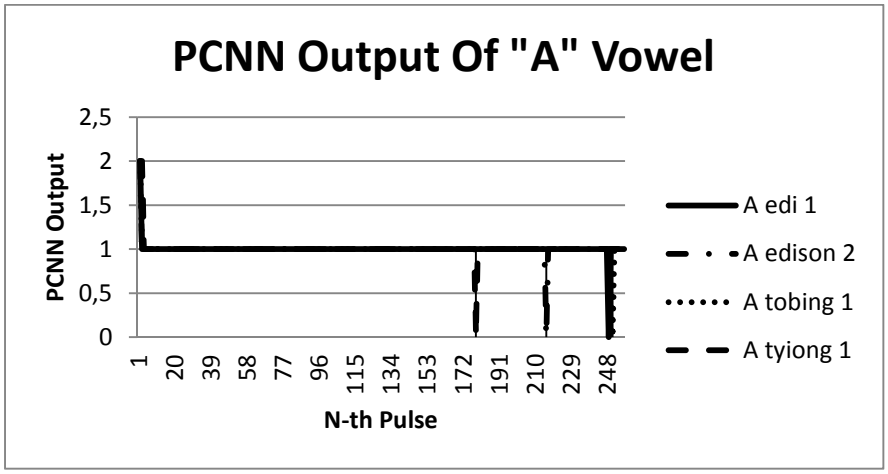

Fig. 3 Output of PCNN for esophageal speech recognition “A”

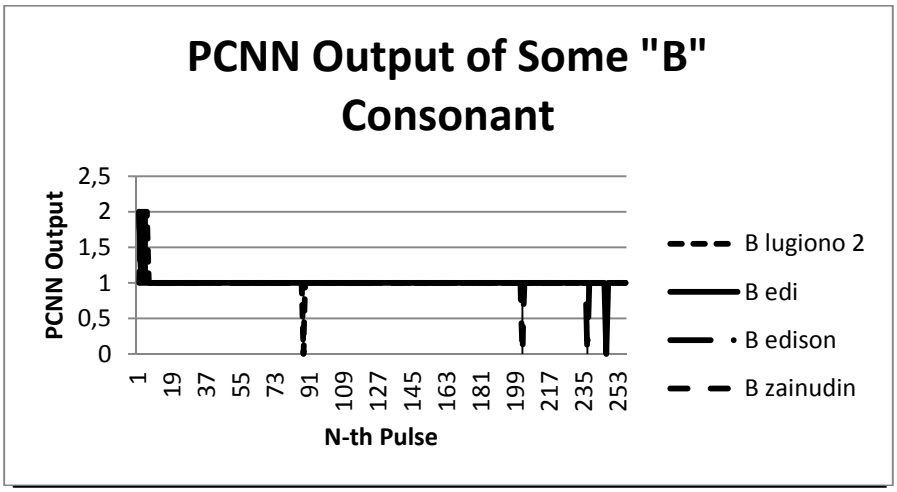

Fig. 4 Output of PCNN for esophageal speech recognition “**B**” Consonant

Furthermore, the output of the PCNN will be accepted by the MLP. MLP has three layers. The number of neurons in each layer: input layer 256, hidden layer 10, and output layer 1.

Furthermore, the system was trained by training set input. In the 363 iteration system met the goal. After the training, the system was tested. The result of training can be seen at table 1.

Table 1 Esophageal Speech Recognition Testing

No	Esophageal Signal	Output	Result of Recognition
1	A edi 1	0.9231	ok
2	A edison 1	0.9340	ok
3	A tobing 1	0.9594	Ok
4	A tyiong 1	0.9432	Ok
5	A zainudin 1	0.9588	Ok
6	A khairul 1	0.9745	Ok
7	A lugiono 1	0.9916	Ok
8	A lugiono 3	0.9726	Ok
9	B lugiono 1	0.0404	Ok
10	B lugiono 2	0.0404	Ok
11	B edi 1	0.0634	Ok
12	B edison 1	0.0914	Ok
13	B Zainudin 1	0.0789	Ok
14	A edi 2	0.6524	Ok
15	A edison 2	0.7738	Ok
16	A tobing 2	0.8814	Ok
17	A tyiong 2	0.8814	Ok
18	A zainudin 2	0.9242	Ok
19	A khairul 2	0.9925	Ok
20	A lugiono 5	0.6657	Ok
21	A lugiono 4	0.0477	Wrong
22	B lugiono 3	0.0292	Ok
23	B tobing 1	0.4354	Ok
24	B tyiong 1	0.0634	Ok

The results shows that the system can recognize 23 from 24 vowel and consonant esophageal speech (95,8 %) . There is only one data that cannot be recognized correctly. This is caused by the poor quality of recording process.

IV. **CONCLUSION**

The result of testing showed that system was running well. It can recognize esophageal speech correctly. Its validity is 91, 2 %. Maybe one data that not recognized correctly is caused by poor quality of recording process.

From the experiment and results we can also conclude that PCNN can be utilized as feature extractor very well. It is simple and gives good results.

REFERENCES

1. Nury Nusdwinuringtyas, Tanpa pita suara berbicara, Blog spot, February, 2010
2. American Cancer Society. Cancer facts and figures-2002
3. Fatchul A, Tri Arief S, Mauridhy Hery, ElectroLarynx, Esopahgus, and Normal Speech Classification using Gradient Discent, Gradient discent with momentum and learning rate, and Levenberg-Marquardt Algorithm, ICGC 2010
4. Fellbaum, K.: Human-Human Communication and Human-Computer, Interaction by Voice. Lecture on the Seminar "Human Aspects of Telecommunications for Disabled and Older People". Donostia (Spain), 11 June 1999
5. http://en.wikipedia.org/wiki/Esophageal_speech, Juli 2010
6. Magdalena Marlin Amanda, application of voice recognition in cryptograph of public key, Tugas Akhir Petra
7. Achmad Basuki, Miftahul Huda, Tria Silvie Amalia, application of voice recognition in musical request, Jurusan Teknik Telekomunikasi Politeknik Elektronika Negeri Surabaya
8. Ajub Ajulian Z., Achmad Hidayatno, Muhammad Widyanto Tri Saksono, application of *voice recognition in* car controller
9. Mohammed Bahoura, Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes, Internationala journal: Computers in Biology and Medicine 39 (2009) 824 -- 843, journal homepage : www.elsevier.com/locate/cbm
10. Taiji Sugiyama, Speech recognition using pulse-coupled neural networks with a radial basis function, ISAROB 2004, Artif Life Robotics (2004) 7:156-159