Electrolarynx Voice Recognition Utilizing Pulse Coupled Neural Network

Fatchul Arifin[1], Tri Arief Sardjono[2], and Mauridhy Hery Purnomo[2]

Abstract—**The laryngectomies patient has no ability to speak normally because their vocal chords have been removed. The easiest option for the patient to speak again is by using electrolarynx speech. This tool is placed on the lower chin. Vibration of the neck while speaking is used to produce sound. Meanwhile, the technology of "voice recognition" has been growing very rapidly. It is expected that the technology of "voice recognition" can also be used by laryngectomies patients who use electrolarynx.This paper describes a system for electrolarynx speech recognition. Two main parts of the system are feature extraction and pattern recognition. The Pulse Coupled Neural Network – PCNN is used to extract the feature and characteristic of electrolarynx speech. Varying of β (one of PCNN parameter) also was conducted. Multi layer perceptron is used to recognize the sound patterns. There are two kinds of recognition conducted in this paper: speech recognition and speaker recognition. The speech recognition recognizes specific speech from every people. Meanwhile, speaker recognition recognizes specific speech from specific person. The system ran well. The "electro-larynx speech recognition" has been tested by recognizing of "A" and "not A" voice. The results showed that the system had 94.4% validation. Meanwhile, the electrolarynx speaker recognition has been tested by recognizing of "saya" voice from some different speakers. The results showed that the system had 92.2% validation. Meanwhile, the best β parameter of PCNN for electrolarynx recognition is 3.**

Keywords—**Electrolarynx speech recognation, Pulse Coupled Neural Network (PCNN), Multi Layer Perceptron (MLP)**

I. Introduction

More than 8900 persons in the United States are diagnosed with laryngeal cancer every year [1]. The average number of laryngeal cancer patients in RSCM is 25 people per 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 radio-active 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 no longer able to 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. Post-surgery of the larynx may cause the patient no longer able to speak as before. There are some ways to make laryngectomies able to talk again. The easiest way is using electrolarynx voice. It is the way to speak using electrolarynx tool. This tool is placed on the lower chin. Vibration of the neck while speaking is used to produce sound. However this sound has a poor quality and it is often not understandable [3].

Meantime research in the speech recognition and its application is growing rapidly. A lot of application of speech recognition was introduced. Some of them are: application of voice recognition in cryptograph of public key published by Magdalena [5], application of voice recognition in musical request published by Achmad [6], and application of voice recognition in car controller published by Ajub [7]. This paper describes how to recognize the electrolarynx speech by utilizing Pulse Code Coupled Network as speech feature extraction.

II. Theories

Electrolarynx speech recognition has two main parts, they are voice extraction and pattern recognition. Voice extraction will extract important parts of the electro-larynx speech characteristics, while pattern recognition is used to identify patterns of electrolarynx speech.

Bahaoura (2009), compare the various methods related to feature extraction and pattern recognition to detect the diseases through the human respiratory sound [8]. He used Fast Fourier Transform (FFT), Linear Predictive Coding (LPC), Wavelet Transform (WT), and MFCC for feature extraction. While in the pattern recognition he used 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 respiratory sound.

In addition to the methods presented by Bahaoura, there is another feature extraction method that is used widely for image processing. This method is Pulse Coupled Neural Network (PCNN).

PCNN is a binary model. Although initially this method is very pupoler just for image processing extraction, now some researchers have developed it for voice recognition. Sugiyama [9] had used PCNN for pattern recognition.

In this paper, electrolarynx speech recognition system consists of Fast Fourier Transform (FFT), Pulse Couple Neural Network (PCNN), and Multi Layer Perceptron (MLP). Block diagram of this processis showed in Fig. 1.

Signal of electrolarynx speech will be converted to the frequency domain by Fast Fourier Transform (FFT). This is important because the frequency domain will give a clearer view to be observed and manipulated than time domain. The output of the Fast Fourier Transform will be fed into the PCNN for getting unique characteristic of

[1] Fatchul Arifin is Student of Electrical Engineering Doctorate Program, FTI, Institut Teknologi Sepuluh Nopember Surabaya, 60111, and Department of Electrical Engineering, FT, Universitas Negeri Yogyakarta, Yogyakarta, 55281, Indonesia. E-mail: fatchul@uny.ac.id.

[2] Tri Arief Sardjono and Mauridhy Hery Purnomo are with Department of Electrical Engineering, FTI, Institut Teknologi Sepuluh Nopember Surabaya, 60111, Indonesia.

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

PCNN is a pair of single layer neural network which is connected laterally, and has two dimensions. In pulsed coupled neuron models, the inputs and outputs are given by short pulse sequences generated over time.

PCNN consist of:

1. Input unit,
2. Linking unit,
3. Pulse Generator unit.

Structure of Pulse Coupled Neural Network in this research can be seen in Fig. 2.

In the input unit there are two parts, namely linking input and feeding input. The feeding input is a primary input from the neuron's receptive area [8]. In this research output signals from FFT is feed to this feeding input. But the signals from FFT output must be normalized first (Equation (1)).

The other hand linking input received feedback signals from output Y (n-1). These signals are biased and then multiplied together (Equation (2)).

Input values F_{ij} and L_{ij} are modulated in linking part of neuron. This process will generate neuron internal activity U_{ij}.

If the internal activity is greater than dynamic threshold, θij, the neuron will generate output pulses. In contrast the output will be zero.

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

F_{ij} = Normalized input (from output of FFT) (1)

$$L_{ij}(n) = L_{ij}(n-1).\ e^{-\alpha L} + V_L.(W*Y(n-1))_{ij} \qquad (2)$$

$$U_{ij}(n) = F_{ij}(n).(1+\beta.L_{ij}(n)) \qquad (3)$$

$$\Theta_{ij}\ (n) = \Theta_{ij}(n-1).\ e^{-\alpha\Theta} + V_\Theta \qquad (4)$$

$$Y_{ij}(n) = \begin{cases} 1 \text{ if } U_{ij} > \Theta_{ij} \\ 0 \text{ otherwise} \end{cases} \qquad (5)$$

where:

F_{ij} : feeding input,
L_{ij} : linking input,
n : iteration number,
W : weight matrix,
* : convolution operation,
Y_{ij} : output of neuron at coordinate i; j,
V_L : linking voltage potential,
α_L and α_F : decayed constants

III. Method and Material

There are two kinds of recognition conducted in this paper; they are speech recognition and speaker recognition. In speech recognition, just a specific voice from some persons will be recognized, on the other hand another voice will not be recognized.

Meanwhile in speaker recognition, specific speech from specific person is recognized. This system cannot recognize another speech from the same person. It cannot recognize the same speech or different speech from another person, either.Firstly, electrolarynx speech recognition conducted. There are 50 sample. They are 28 ”A” vowel and 22 ”not” from different electrolarynx speaker. Electrolarynx speech samples from data base were divided into training sets and testing sets.

Secondly “Electrolarynx speaker” recognition was conducted. There are 28 “saya” electrolarynx voice from different speaker.

IV. Experiments and Results

A. Electro Larynx Speech Recognition

Recorded electrolarynx 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 twice as high as input frequency. Thus, 44100 Hz sampling has fulfilled the nyquist criterion. Further, sampled signal will be converted from time domain into frequency domain utilizing Fast Fourier Transform (FFT). In this paper it used FFT 512 point. Because the FFT is symmetric, the FFT output is taken only half which is 256 data. All of electrolarynx voice signal (for training or testing in MLP) are processed by this FFT. One of the FFT output signal can be seen at Fig. 3.

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);

Some output signals of PCNN training set data can be seen at Fig. 4.

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

The system was trained by training set input. In the 359 iteration the system met the goal. After the training, the system was tested. The result shows that the system can recognize 47 from 50 sample of electrolarynx speech (94 %). There is only 3 data that cannot be recognized correctly.

B. Electrolarynx Speaker Recognition

In this session it will be shown that system also can recognize “Electrolarynx speaker” correctly. There are 28 “saya” of electrolarynx voice from different speaker. These Electrolarynx voice samples were divided into training and testing sets.

Furthermore, these signals were processed like at electrolarynx speech recognition processing in above. The PCNN parameters that used were the same as before. Some PCNN Output of “saya” electrolarynx speech that used for training set can be seen in Fig. 5.

Then, the system was trained by training set. At the 250th iteration, the system met the goal. After training process, the system will be tested. The testing result shows that the system able to recognize 26 from 28 sample of electrolarynx speech. It means that it has validation 92,2 %. There is only 2 data that cannot be recognized correctly.

C. β Parameter of PCNN

As mentioned above, the output of PCCN is greatly affected by its parameters. The most significant parameter is β which is the linking parameter. With different β, it will be got different weight in linking channel internal activity. In the end, it will affect of PCCN output. In the experiments above, it used β = 3.

In this paper, it will be shown what happens if the Fig. 6 shows the graph of PCCN Output for some variation of β.

Furthermore, these PCNN Outputs (with differences β) were used for electrolarynx speaker recognition like above. The result of the test can be seen in the Table 1."True" means it is recognized, and false means it is not be recognized. From table above, it can be seen that PCCN with β coefficient 3 gives the best result with validation up to 92.2%.

The decrease of β value causes the decrease of validation. Meanwhile, the increase of β value also decreases its validation. Therefore, the best value of β in this research is 3.

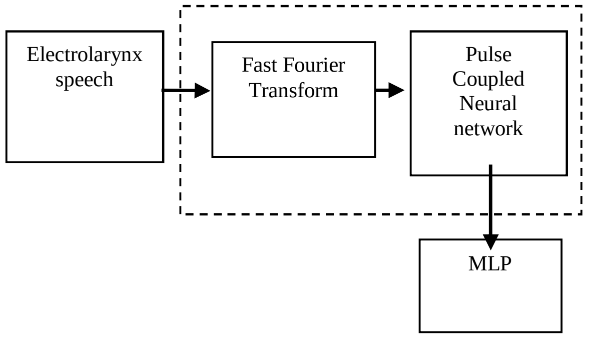

Fig. 1. Electrolarynx speech recognition system

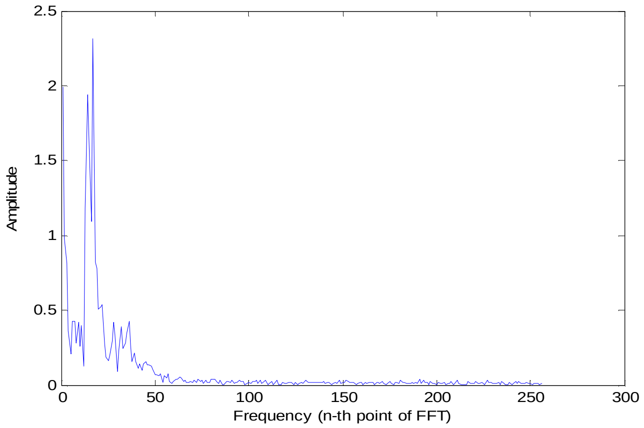

Fig. 3. FFT signal output of Electrolarynx voice

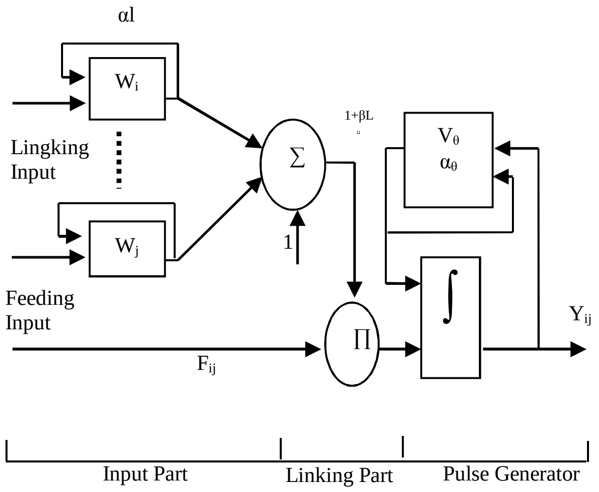

Fig. 2. Structure of PCNN

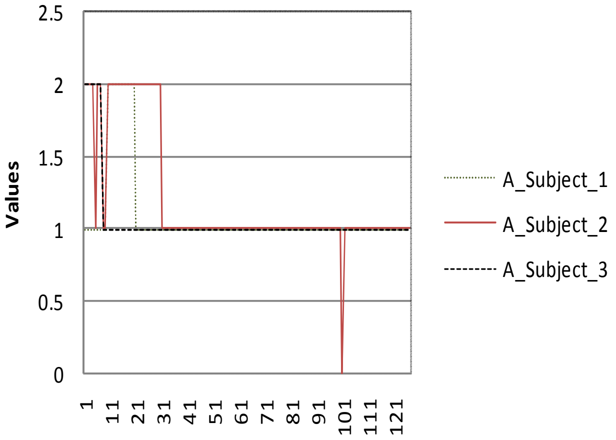

Fig. 4. Output of PCNN of "A" vowel electrolarynx speech

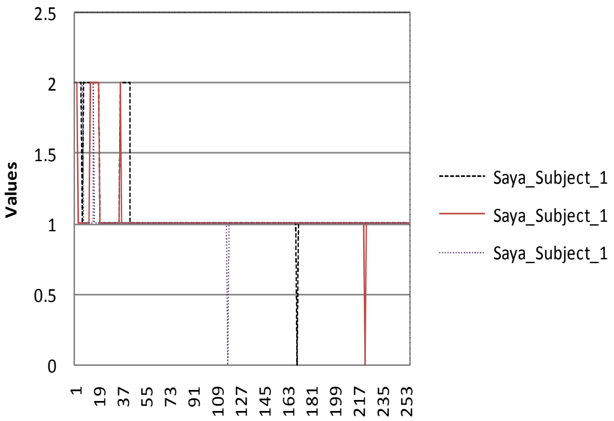

Fig. 5. Some PCNN output of “saya” electrolarynx speech

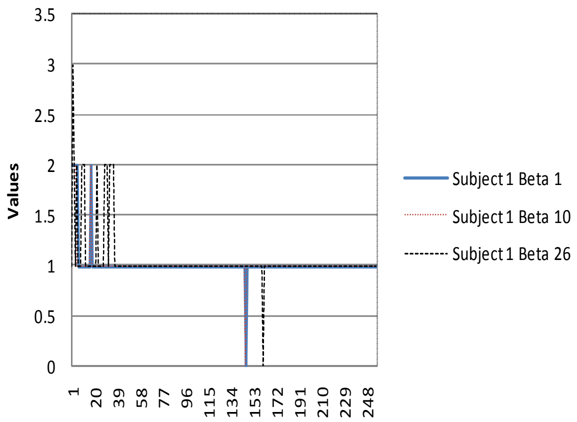

Fig. 6. PCNN output of “saya” electrolarynx speech with variation of Beta (β)

TABLE 1.
COMPARISON OF ELECTRO LARYNX SPEAKER RECOGNITION OUTPUT WITH DIFFERENCES PCNN’S β

Voice		Beta (β)					
		0.1	1	3	10	26	40
1.	Saya_subject1_1	True	True	True	True	True	True
2.	Saya_subject1_2	True	True	True	False	False	True
3.	Saya_subject1_3	True	True	True	True	True	True
4.	Saya_subject1_4	True	True	True	True	True	True
5.	Saya_subject1_5	True	True	True	True	True	True
6.	Saya_subject1_6	True	True	True	True	True	True
7.	Saya_subject2_1	False	False	True	True	True	True
8.	Saya_subject2_2	True	True	True	True	True	True
9.	Saya_subject3_1	True	True	True	True	True	True
10.	Saya_subject3_2	True	True	True	True	True	True
11.	Saya_subject4_1	True	True	True	True	True	True
12.	Saya_subject4_2	True	True	True	True	True	True
13.	Saya_subject4_5	True	True	True	True	True	True
14.	Saya_subject1_7	False	False	True	False	True	False
15.	Saya_subject1_8	True	True	False	True	True	True
16.	Saya_subject1_9	False	False	True	True	False	True
17.	Saya_subject3_3	True	True	True	True	True	True
18.	Saya_subject3_4	False	False	True	True	True	False
19.	Saya_subject3_5	True	True	True	False	False	False
20.	Saya_subject3_6	True	True	True	True	False	False
21.	Saya_subject2_3	False	True	True	True	False	False
22.	Saya_subject2_4	False	False	True	False	False	False
23.	Saya_subject2_5	False	False	True	True	True	False
24.	Saya_subject2_6	False	False	True	True	False	False
25.	Saya_subject2_7	True	True	False	False	True	True
26.	Saya_subject4_3	False	False	True	True	False	False
27.	Saya_subject4_6	False	False	True	True	True	True
28.	Saya_subject4_7	False	True	True	False	False	False
Percentage of Correctness (%)		60.7	67.9	92	78.6	67.9	64.2

Where true: It can be recognized correctly, false: It cannot be recognized correctly

V. CONCLUSION

The result of testing showed that the system ran well. The "electrolarynx speech recognition" has been tested by recognizing “A” and "not A" voice. The results showed that the system had 94.4% validation. Meanwhile, the electrolarynx speaker recognition has been tested by recognizing “saya” voice from some different speakers. The results showed that the system had 92.2% validation.

One of the important PCNN parameters was β. By tuning β in specific values, it reached the best output. In this paper the best β value for Electrolarynx speaker recognition is 3.

REFERENCES

[1] Alvin G. Wee, BDS, MS, a Lisa A. Wee, 2004, “The use of an intraoral electrolarynx for an edentulous patient: A clinical report, Ohio State University, Columbus, Ohio; University of Toronto, Toronto, Vol. 91, pp. 521.

[2] http://www.wikimu.com/News/Print.aspx?id=11467. Data medis departemen rehabilitasi medis, RSCM, February 2010

[3] A. Subject2chul, S. Tri Arief, Mauridhy Hery, 2010, “ElectroLarynx, Esopahgus, and normal speech classification using gradient discent, gradient discent with momentum and learning rate, and levenberg-marquardt algorithm, *ICGC*.

[4] K. Fellbaum, 1999, “Human-human communication and human-computer, interaction by voice”. *Human Aspects of Telecommunications for Disabled and Older People*, Donostia, Spain.

[5] Magdalena Marlin Amanda, application of voice recognition in cryptograph of public key, Tugas Informatika, ITB, 2009

[6] Achmad Basuki, Miftahul Huda, Tria Silvie Amalia, 2006, “Application of voice recognition in musical request,”

Proceeding of IES (Industrial Elektronic Seminar) Politeknik Elektronika Negeri Surabaya-ITS.

[7] Ajub Ajulian Z., Achmad Hidayatno, Muhammad Widyanto Tri Saksono, application of voice recognition in car controller, Majalah Transmisi, Teknik Elektro Undip, Jilid 10, Nomor 1, Maret 2008

[8] Mohammed Bahoura, Pattern,2009, "Recognition methods applied to respiratory sounds classification into normal and wheeze classes", *Internationala journal: Computers in Biology and Medicine*, Vol. 39 pp. 824-843. www.elsevier.com/locate/cbm

[9] Taiji Sugiyama, 2004,"Speech recognition using pulse-coupled neural networks with a radial basis function", *ISAROB 2004*, Vol. 7, pp. 156-159.