International Conference
Arifin, F., Sardjono, T.A., Purnomo, M.H., Esophageal Speech Recognition Utilizing Pulsed Coupled Neural Network, International Seminary BME Days (Biomedical Engineering), 2010, Surabaya

Sertifikat
Letter of Acceptance

Dear Fatchul Ariffin

We hereby inform you that your paper with

Paper ID: 014
Title: Esophageal speech Recognition Utilizing Pulse Coupled Neural Network

has passed the review stage and is ACCEPTED WITH CORRECTION(S)

Reviewer's Notes:
1. Perhatikan lagi tulisannya
2. Kesimpulan bisa ditambahkan

The payment and registration should be completed before September 1, 2010. Send your scanned payment transcript via e-mail, with your paper ID as the file name (example: 001-Proof of Payment). For more complete informations please go to our home page: http://bme-days.ieteeexpo.org

Congratulations for your paper acceptance. We are very pleased to present your ideas on BME Days 2010.

Best regards,

Prof. Dr. Ir. Mensiti Jary Pamono, M.Eng.
General Chairman of BME Days 2010
Cover
Susunan editor

PROCEEDINGS OF THE 6TH INTERNATIONAL CONFERENCE ON BIOMEDICAL ENGINEERING,
BME DAYS 2010

EDITOR-IN-CHIEF
I Ketut Eddy Purnama, ST., MT., Ph.D
E-mail : ketut@ee.its.ac.id

EDITORIAL COORDINATORS
Achmad Arifin, ST., M.Eng., Ph.D
E-mail : arifin@ee.its.ac.id

Dr. Tri Arief Sardjono, ST., MT.
E-mail : t.a.sardjono@ee.its.ac.id

I Ketut Eddy Purnama, ST., MT., Ph.D
E-mail : ketut@ee.its.ac.id

Dr. Mohamad Rivi, ST., MT.
E-mail : mahamad_rivi@ee.its.ac.id

EDITORIAL ADVISORY BOARD
Prof. Dr. Ir. Mauritdi Hery Purnomo, M.Eng.
E-mail: bery@ee.its.ac.id
COMMITTEE OF BME DAYS 2010

General Chair:
Prof. Dr. Ir. Maurichsi Heri Purnomo (ITS, Indonesia)

Advisory Committee:
Prof. Dr. Ir. M. Nuh, DEA (ITS, Minister of Education)
Prof. Ir. Priyo Surobo, MS., Ph.D (Rector of ITS, Indonesia)
Dr. dr. Nasronuddin Sp,FD, K-FTL (Unair)
Prof. Dr. dr. Kuntamun, MS., Sp.MK (Unair)
Prof. Dr. Bart Verkertoe (Rijksuniversitit Groningen, Netherlands)

Technical program committee:
Dr. I Kerus Eddy Purnama (ITS, Indonesia)
Dr. Prihartini Widyantari, drg. M.Kes (TD-Unair, Indonesia)
Dr. Tri Arief Surjono (ITS, Indonesia)
Dr. Ahmad Arifin (ITS, Indonesia)
Dr. Muhammad Riza (ITS, Indonesia)
Dr. Udjo Purwanto (ITS, Indonesia)
Dr. Agus Zainal Arifin (ITS, Indonesia)
Dr. Dadi Pramadiharta (ITS, Indonesia)
Dr. rer. nat. Aulia M.T. Nasution (ITS, Indonesia)

Organizing Committee
Atar Fuady Babego
Pratama Ari Baskoro
Riyan Sujaitmiko
Kevin Nyoman Putra
Arto Nugroho
Anci Nur Ariew Wibowo
Gunarsih Wicaksoro
Median HP

Niko Permana R.W
Febi Ardianto
Rizly Pratama Putra
Radka Henry Wijaya
Fery Efendi Hermawan
Siti Asyah
Dial Kusumawati
Alfisa Zuzaida

Prof. Dr. Tati Mengko (ITB)
Prof. Dr. Widowati Siswunindhurjo (UGM)
Prof. Dr. Tresna P. Soemardjo (UI)
Prof. Dr. Rosidin Murti (Unhas)
Prof. Dr. med. Furuhito, MD. (Unair)

Dr. Son Kuswadi (ITS, Indonesia)
Dr. Dhany Ariefananto (ITS, Indonesia)
Dr. Suprahto (ITS, Indonesia)
Dr. M. Zainuri (ITS, Indonesia)
Dr. Darminto (ITS, Indonesia)
Syamseul Arifin MT. (ITS, Indonesia)
Syuanto MT. (ITS, Indonesia)
Diah Puspito W. (ITS, Indonesia)

Ivan Karsten
M. Nur Rohman
Harno Pratomo
M. Ikhean
Debby Pratiwi Melyesti T.
LIST OF PAPERS

001
Towards Quantitative, Low-Cost, Webcam-Based Microscopy for Medical Diagnostics
Maden Darmawan, Anir Faisal, Oea O.F. Parikesit

002
Bio-computational Analysis of Rheumatoid Arthritis of the Wrist
Mold Nazri Bajuri, Mohammed Rafiq Abdul Kadir

003
Cry Spectrography as a Diagnostic Tool of Acute Pain in Infants (A Preliminary Study)
Elaineus Hanindito, Nincy Margarita Relatta, Eddy Rahardjo, Mauridi Hery P.

004
Color Retinal Image Compression Using Adaptive Vector Quantization
Agung W. Setiawan, Andriyani B. Suksumono, Tati R. Mengko, Octip S. Santoso

005
Spectral Analysis of EEG Signals during Listening Music and Math Task Conditions
Kenalasari, Mauridi Hery Purnomo

006
QR Code Application In Wordpress Based For Medical Learning With Online/Offline Scenario With RSS Synchronization
Ahmad Sirejuddin, Andra Ahmad, Admad Affandi

008
Radiation Exposure Measurement Outside LINAC Radiotherapy Basuk of Dr. Sardjito Hospital, Yogyakarta
Yuwono Marta Dinna, Masedjdi Tj, Thomas Sri Widodo

009
Emotion Recognition Using Physiological Analysis and Pattern Matching of Face Sequences from Video
Arkondo Chris, Lina

010
Implementation of Centralised Electronic Health Record Data Center based on OpenEHR in Indonesia
Alasanai Atmopawino, Lovinta Happy Atriwarni, Paul Gunawan Haryanto, Chitra Hapsari Ayuningtyas, Putri Sapruwati

012
A General Mapping Tool for Exchanging Electronic Health Records
Lovinta Happy Atriwarni, Alisanai Atmopawino, Paul Gunawan Heryanto
013
Pulse Waveform Records of Type 2 Diabetic Nephropathy Disease Patients at Radial Artery Chi Area
Enni Yudaningtyas, Harjono Achmat, Rasjod Indra, Waru Djuriatno

014
Esophageal Speech Recognition Utilizing Pulse Coupled Neural Network
Fatchul Arifin, Tri Arief Sardjono, Maudidi Hery Pumomo

016
Rician Denoising from MRI Images via ROF Model
Agung Alfiansyah

017
Smart Acquisitions ECG Data Using 3G with Mobile Telemedicine
Muhammad Ashar

019
Heartbeat Biomodeling with Pressure Change Approach for CABbaGe Simulator
Arieek Ida Wuryansari, Zakia Lutfiyyani

020
Design And Implementation of Vessel Cutting Simulation Using Simulation Open Framework Architecture (SOFA)
Arieek Ida Wuryansari, Gina Indriati

021
Development of ST-Segment Interpretation Software for Myocardial Ischemia and Infarction Using Artificial Neural Network Multilayer Preceptor
Aulia Arif Iskandar, Judejono Kartijo, Tuti L. R. Mengko

022
Two Video Synchronization And Three Dimensional Reconstruction for Human Gait Viewing
Amar Vijai N., Tuti L. R. Mengko, G. Apriantono

023
Effect of Cold Deformation and Sandblasting on Hardness, Microstructure and Corrosion Resistance of AISI316L for Medical Implant
Teguh Dwi Widodo, Suyitno

024
Stress Transfer of Different Abutment-Implant Connection of Dental Implant
Aisyah Ahmad Shafi, Mohammed Iafiq Abdul Kadir, Shamsul Sulaiman
014 Esophageal Speech Recognition Utilizing Pulse Coupled Neural Network

Fatchul Artini¹, Iri Arief Sardjono¹, Maurydhdy Hery Purnomo²
1. Jurusan Teknik Elektro Universitas Negeri Yogyakarta, fatchul@uns.ac.id
2. Jurusan Teknik Elektro Institut Teknologi Surabaya, email: i.s.sardjono@uns.ac.id, hery@itek.it.ac.id

Abstract - The laryngectomy patients have no ability to speak normally because their vocal chords have been removed. The simplest option for the patient to speak 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 part 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 characteristics 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 91.8%. It is also can be known that PCNN can be utilized as feature extractor very well.

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

I. INTRODUCTION

The average number of laryngeal cancer patients in RSCM ± 25 people per year[1]. More than 8,900 persons in the United States are diagnosed with laryngeal cancer every year [2]. The exact causes and of the larynx until now is unknown, but it is found some things that are closely related to the occurrence of laryngeal malignancies: cigarette smoking, alcohol, and radioactive rays.

Laryngectomy is a type of surgery needed to make a hole (stoma) - a particular part of body. Laryngectomy is an example of Osteomy. It is an operation performed on patients with cancer of the larynx (thyroid) 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 vibration of the tongue, the vocal throat (epiglottis) with the vocal cords and articulation created 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 anywhere.

Several ways to make laryngectomy can talk again have been developed. The easiest way is esophageal speech. Esophageal speech is a voice generated without the oscillation of the vocal cords. The voice is produced by releasing gases through the esophagus in a manner similar to burping, to create speech. The esophageal function to esophageal speech is in much the same manner as the vocal cords in laryngeal speech, oscillating quickly to create distinct speech sounds. Esophageal speech is speaking by expiration [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 maqbool [6], application of voice recognition in the signal request by amin dan banti [7], application of voice recognition in car controller by ajoy aijias [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 Coupled Neural 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 characteristics of the esophageal speech, while pattern recognition is utilized to identify patterns voices.

In the year 2009, Mohammad Bahauddin compare the various methods related to feature extraction and pattern recognition for detection of diseases through the human respiratory sound [9]. In the feature extraction he utilized some different methods: the Fast Fourier Transform (FFT), Linear Predictive Coding (LPC), Wavelet Transform (WT), and Mel-frequence cepstral coefficients (MFCC). He also utilized some different methods for pattern recognition process, they are Vector quantization (VQ), Gaussian Mixture Hidden (HMM) and Artificial Neural Network (ANN). According to Bahauddin, The combination between MFCC and HMM is the best methods related to the respiratory sound.

Beyond what is presented by bahauddin, 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 popular for image processing extraction, but now some researcher develop it for voice recognition. Taiki Sugiyama utilized PCNN as pattern recognition unit [10].

Esophageal speech recognition system which is proposed in this paper consists of Fast Fourier Transform (FFT), Pulse Coupled Neural Network (PCNN), and Multi Layer Perceptron (MLP). Block diagram of this process 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 multilayer 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 consists 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:

\[ u_d(n) = u_d(n-1) + e^{W_L}U \]  
(1)

\[ u_i(n) = u_i(n-1) + u_d(n) \]  
(2)

\[ U = 0.6 \times (U_l > U_0) \]  
(3)

1 if \( U_l > U_0 \)

0 otherwise

Where:

- \( W \) : feeding input,
- \( I \) : linking input,
- \( n \) : iteration number,
- \( W \) and \( M \) : weights matrix,
- \( \ast \) : convolution operation,
- \( Y_{ij} \) : output of neuron at coordinate \( ij \),
- \( V_L \), \( V_F \) : voltage potential,
- \( \alpha \) and \( \gamma \) : decayed constants

Single signals of the linking input are biased and multiplied together. Input values \( W \) and \( I \) are modulated in linking part of neuron. These process will generate neuron internal activity \( U_i \). If the internal activity is greater than dynamic threshold \( U_0 \), 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 matrices. Each of these binary matrices has the same dimension as input matrix. The sum of 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

Fig. 3 Output of PCNN for esophageal speech recognition "A".

Fig. 4 Output of PCNN for esophageal speech recognition "B" Consonant.
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 larynges on the voice. Speech samples from data base were divided in to training and testing sets.

Recorded esophageal speech was sampled with sampling frequency 64.100 Hz and 16 bit 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:

\[ f_s \geq 2f_N \]

\[ f_N = f_{Max} \]

Sampling frequency must be equal to or greater than twice input frequency. Thus, 64.100 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 156 data.

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

- \( a_1 = 1 \)
- \( a_2 = 0.2 \)
- \( b = 5 \)
- \( V_i = 1.00 \)
- \( V = 0.2 \)
- \( L = \text{zeros}(p,q); \)
- \( U = \text{zeros}(p,q); \)
- \( V = \text{zeros}(p,q); \)
- \( Y_0 = \text{zeros}(p,q); \)
- \( \Theta = \text{zeros}(p,q); \)

The same 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.

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 246, hidden layer 10, and output layer 1.

Furthermore, the systems were 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>
<thead>
<tr>
<th>No</th>
<th>Esophageal Signal</th>
<th>Output</th>
<th>Result of Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A eed 1</td>
<td>0.9231</td>
<td>ok</td>
</tr>
<tr>
<td>2</td>
<td>A ediot 1</td>
<td>0.9340</td>
<td>ok</td>
</tr>
<tr>
<td>3</td>
<td>A eing 1</td>
<td>0.9594</td>
<td>Ok</td>
</tr>
<tr>
<td>4</td>
<td>A eeing 1</td>
<td>0.9432</td>
<td>Ok</td>
</tr>
<tr>
<td>5</td>
<td>A eainle 1</td>
<td>0.9588</td>
<td>Ok</td>
</tr>
<tr>
<td>6</td>
<td>A eahed 1</td>
<td>0.9745</td>
<td>Ok</td>
</tr>
<tr>
<td>7</td>
<td>A hution 1</td>
<td>0.9016</td>
<td>Ok</td>
</tr>
<tr>
<td>8</td>
<td>A hution 2</td>
<td>0.9705</td>
<td>Ok</td>
</tr>
<tr>
<td>9</td>
<td>B higation 1</td>
<td>0.0059</td>
<td>Ok</td>
</tr>
<tr>
<td>10</td>
<td>B higation 2</td>
<td>0.0604</td>
<td>Ok</td>
</tr>
<tr>
<td>11</td>
<td>B eed 1</td>
<td>0.0634</td>
<td>Ok</td>
</tr>
<tr>
<td>12</td>
<td>B eed 2</td>
<td>0.0014</td>
<td>Ok</td>
</tr>
<tr>
<td>13</td>
<td>B Zainle 1</td>
<td>0.0782</td>
<td>Ok</td>
</tr>
<tr>
<td>14</td>
<td>A ede 2</td>
<td>0.6324</td>
<td>Ok</td>
</tr>
<tr>
<td>15</td>
<td>A ede 2</td>
<td>0.7738</td>
<td>Ok</td>
</tr>
<tr>
<td>16</td>
<td>A ede 2</td>
<td>0.8214</td>
<td>Ok</td>
</tr>
<tr>
<td>17</td>
<td>A eing 2</td>
<td>0.8814</td>
<td>Ok</td>
</tr>
<tr>
<td>18</td>
<td>A eing 2</td>
<td>0.9242</td>
<td>Ok</td>
</tr>
<tr>
<td>19</td>
<td>A bialled 2</td>
<td>0.9932</td>
<td>Ok</td>
</tr>
<tr>
<td>20</td>
<td>A bialled 5</td>
<td>0.6567</td>
<td>Ok</td>
</tr>
<tr>
<td>21</td>
<td>A higation 4</td>
<td>0.0477</td>
<td>Wrong</td>
</tr>
<tr>
<td>22</td>
<td>B higation 3</td>
<td>0.0992</td>
<td>Ok</td>
</tr>
<tr>
<td>23</td>
<td>B tring 1</td>
<td>0.4354</td>
<td>Ok</td>
</tr>
<tr>
<td>24</td>
<td>B tring 1</td>
<td>0.0634</td>
<td>Ok</td>
</tr>
</tbody>
</table>

The results show 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.

V. 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


[8] Ajiub Ajiadien Z, Achmad Hidayanto, Muhammad Widyanoto Tri Saksmono, application of voice recognition in car controller
