Feature Extraction of Facial Electromyography (EMG) Signal for Aceh Languages Speech using Discrete Wavelet Transform (DWT)

Darma Setiawan Putra\textsuperscript{a,1,*}, Yurl Umbu Woza Weru\textsuperscript{b,2}

\textsuperscript{a,b} Informatics Engineering, Politeknik Aceh Selatan, Merdeka Street, Tapaktuan, 23711, Indonesia
\textsuperscript{1}darma.poltas@gmail.com\textsuperscript{a}; \textsuperscript{2}yurlweru@gmail.com

\* corresponding author

I. Introduction

The development of science and technology has significantly influenced in signal processing technology. The human has electromyography (EMG) signal in the arm, leg, and facial muscle. The EMG signal in the face is called the facial electromyography (FEMG) signal. The FEMG is usually acquired by placing an electrode surface on the skin around the facial articular muscle. Three types of muscles in this study are the masseter, rigorous and depressor muscle. This study aims to extract and analyze the features in the FEMG signal. The extraction method is the discrete wavelet transform (DWT). The type of wavelet transform is Daubechies2 with level 5. After extraction and analysis of FEMG signals, the FEMG signal pattern for each spoken word indicated by differences in the approximation and detail coefficient of the FEMG signal. In addition, the level of difference in the FEMG signal pattern is also indicated by the histogram of the approximation coefficient of the FEMG signal. Thus, the discrete wavelet transform method can be used as one of the methods for extracting the FEMG signal feature in a human facial electromyography (FEMG) signal.

Keywords: Discrete wavelet transform, Speech recognition, Feature extraction.
II. Related Studies

Some previous studies about feature extraction methods based on facial electromyography (FEMG) signals have been carried out. The feature extraction method using a discrete wavelet transform for the application not only signal extraction but also image extraction [11]. Previous research by [12] has conducted a study to extract the EMG signal features for Spanish syllable pronunciation using the Fast Fourier Transform (FFT) method. The resulted features are 20 components namely root mean square, average amplitude, maximum amplitude, kurtosis, mel-frequency cepstral coefficient as many as 13 values, mean absolute value, zero crossing. The syllables used are 5 units, namely vowel, labials, dentals, palatals, velars, and alveolar. The next research conducted by [8] to extract the EMG signals for Thai pronunciation using the time domain, frequency domain, and time-frequency domain methods. The features obtained are 68 units with the top 5 features that have the selected RES Index value. The spoken syllables are 21 Thai words with 5 tones for each word. The next research conducted by [7] to extract the EMG signals for Bangli spelling pronunciation using 4 types of methods, namely time-based, frequency based, entropy-based and time-frequency based. The features obtained are 57 units. The syllables are as many as 11 Bangli letters. In this study, we performed facial EMG extraction using discrete wavelet transform to pronounce five syllables in the Acehnese language.

III. Method

A. Flowchart research

The stages of this study as shown by the flow diagram in Fig. 1 where the approximate coefficient component and wavelet detail coefficient are obtained by decomposing the wavelet signal on each existing facial EMG data.

B. Data collection

At this stage, the need for the tools needed to retrieve FEMG signals is determined. The tools and equipment are the surface electrode and FEMG sensor as shown in Fig. 2, Arduino Uno mini computers, cables, and laptop.

![Flowchart of Research](image)
The FEMG signal data is recorded and acquired by attaching an EMG surface electrodes sensor to the facial muscles namely the masseter muscle, risorius, and depressor. The electrode placement point for FEMG data retrieval as shown by Fig. 3. Electrodes in the masseter muscle as the reference electrode.

The recording of EMG signal uses the Cool Term software. Each word pronunciation is given time for initializing and ending. The initialization and termination times are given for 5 seconds. Initialization time is to normalize the condition of the muscles in the face before speaking the word. While the termination time is to normalize the condition of the muscles in the face after a speech the word. The number of participants involved in taking this data is one person. The five Acehnese vocabulary words were spoken as shown in Table 1.

<table>
<thead>
<tr>
<th>No</th>
<th>English Language Vocabulary</th>
<th>Aceh Language Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Where</td>
<td>Ho neuk jak</td>
</tr>
<tr>
<td>2.</td>
<td>Go</td>
<td>Ja’</td>
</tr>
<tr>
<td>3.</td>
<td>Go home</td>
<td>Ja’ woe</td>
</tr>
<tr>
<td>4.</td>
<td>Come from</td>
<td>Panè</td>
</tr>
<tr>
<td>5.</td>
<td>Who</td>
<td>Soe</td>
</tr>
</tbody>
</table>

Darma Setiawan Putra et al (Feature Extraction of Facial Electromyography (EMG) Signal for Aceh Languages...
C. Data Processing

After the signal acquisition, the FEMG signal was processed and analyzed using Matlab software. The FEMG signal data was analyzed using the wavelet transform method. The FEMG signal was decomposed first by using two filters namely low pass filter (LPF) and high pass filter (HPF). After decomposition, the FEMG signal was produced two components, namely the approximation coefficient and the detail coefficient. The illustration of the decomposition process is shown in Fig. 4.

![Wavelet Transform Decomposition in EMG Signal](image)

Fig. 4. Illustration of wavelet transform decomposition in EMG signal

Fig. 4 explains that a high pass filter produces a detail coefficient while a low pass filter produces an approximation coefficient. Down sampling is only done at the output of the low pass filter. This down sampling process was be repeated using a high pass filter and low pass filter. This research uses the wavelet transform with type Daubechies2 to level 5.

IV. Results and Discussion

Using the discrete wavelet transform method as described in section III above, a facial EMG signal is obtained for pronouncing five Acehnese vocabulary words as shown in Table 2.

### Table 2. Facial EMG signal at the time of the speech

<table>
<thead>
<tr>
<th>No</th>
<th>Vocabulary</th>
<th>Facial EMG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Ho neuk jak</em></td>
<td><img src="image" alt="Graph of EMG Signal" /></td>
</tr>
</tbody>
</table>

*Darma Setiawan Putra et al (Feature Extraction of Facial Electromyography (EMG) Signal for Aceh Languages...)*
The form of the FEMG signal in the approximation coefficient and the detail coefficient wavelet for the daubechies2 level 5 for each vocabulary word as shown in Table 3.
Table 3. The EMG facial signals in the approximate coefficients and detail coefficients wavelet.

<table>
<thead>
<tr>
<th>No</th>
<th>Aceh Vocabulary and Wavelet Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ho neuk jak :</td>
</tr>
<tr>
<td>2.</td>
<td>Ja' :</td>
</tr>
<tr>
<td>3.</td>
<td>Ja' woe :</td>
</tr>
</tbody>
</table>

---

**Darma Setiawan Putra et.al (Feature Extraction of Facial Electromyography (EMG) Signal for Aceh Languages...)**
Table 3 shows that the form of the wavelet signal for the approximate coefficients of level 5 is different for each word spoken. Likewise, the shape of the wavelet signal for the Daubechies2 detail coefficient from level 1 to level 5 is also different for each word spoken. This indicates that the signal feature of the FEMG signal has a difference for each spoken word.

The histogram for the approximate coefficient level 1 of the FEMG signal for each word was pronounced as shown in Fig. 5 to Fig. 9.

Fig. 5. Histogram of the approximate coefficient for ho neuk jak

Note: s: EMG signal, a: approximate coefficient, d: detail coefficient
Fig. 6. Histogram of the approximate coefficient for $ja'$

Fig. 7. Histogram of the approximate coefficient for $ja'\ woe$

Fig. 8. Histogram of the approximate coefficient for $panè$
Fig. 5 to Fig. 9 explain that the histogram of approximate coefficient level 1 for each vocabulary word differs the level bar value for each figure according to the word spoken. Fig. 5 shows the maximum level bar with a value of 0.16 in the range of 300 – 310. Fig. 6 shows the maximum level bar with a value of 0.12 at 295 – 300. Fig. 7 shows the maximum level bar with a value of 0.09 in the range of 295 – 300 and 300 – 305. Fig. 8 shows the level bar with a value of 0.37 in the range of 260 – 280. Fig. 9 shows the maximum level bar with a value of 0.14 in the range of 285 – 290. It shows that the FEMG signal for each word pronunciation differently produces a FEMG signal pattern in the risorius, depressor and masseter muscles one another. This EMG signal pattern shows that the humans have different articulation of facial muscle movements for each word spoken so that the resulting FEMG signal features are different from one another. Thus FEMG signal feature can be used to distinguish one another FEMG signal for each word pronunciation.

V. Conclusion

Facial electromyography signals are signal produced because of the contraction of muscles in the human face. The FEMG signal pattern can be known by generating the signal feature. The signal feature is obtained by extracting and analyzing FEMG signals. Extract and analyze of signal feature use the discrete wavelet transform method that aims to determine the level of difference between one another signal at the time of word pronunciation through the approximate coefficient and detail coefficient of the FEMG signal. The results show that the histogram maximum level bar with a value of 0.16 for ho neuk jak, 0.12 for ja’, 0.09 for ja’ woe, 0.37 for panè, 0.14 for Soe.

Acknowledgment


References


