Japanese Vowels Recognition Using Linear Discriminant Analysis and Surface Electromyogram Measured with Bipolar Dry Type Sensors

Ryosuke Takabatake*, Shin-ichi Ito*, Momoyo Ito⁺ and Minoru Fukumi⁺

Tokushima University
2-1, Minami-Josanjima, Tokushima, 770-8506, Japan

* {c501637033, s.ito}@tokushima-u.ac.jp
+{momoito, fukumi}@is.tokushima-u.ac.jp

Abstract

This paper proposes a Japanese vowels recognition method using surface electromyogram (EMG). First, 3 sensors are used to measure surface EMG data at orbicularis oris muscle, zygomatic muscle and depressor angle oris muscle. Next, Fast Fourier Transform (FFT) is applied to all measurement data to calculate power spectra. Linear Discriminant Analysis (LDA) is then used for power spectra of 3 channels and reduce their dimension to 4. Finally, the result of LDA is recognized by Support Vector Machine (SVM). In experiments, it is assumed that mounting sensors to face, measuring EMG, and demounting them are 1 trial. A subject utters 5 Japanese vowels 3 times. Among 3 trials data, 2 trials data are used to make templates and the remaining are used for test. The subject is a man in twenties. As a result, we obtained 62.3% average recognition accuracy. This result shows the proposed method is better about 2 times than the previous method ⁽¹⁾.

Keywords: EMG, vowels recognition, LDA, SVM.

1. Introduction

Voice communication is the most essential and important means of human communication. However, in late years there are troubles where noisy utterance is a nuisance to people, and it is difficult for some people (vocalization dysfunction) to utter by acquired diseases, accidents, and so on.

Patients of vocalization dysfunction cannot talk by themselves. Today, they use alternative utterance methods,

using Electrolarynx, esophageal speech and so on. However, these are basically inconvenient. Therefore, voice recognition and utterance assistance have been studied in recent years. For example, there are methods to recognize utterance with an image of lips shape ⁽²⁾ and using surface EMG signals ⁽³⁾⁽⁴⁾. These methods don't need actual utterance. Therefore, it is thought these are valid to the patients.

In this paper, we focus on a method using EMG signals. This method attaches small electrodes on a line of lips neighborhood and measures EMG signals. After that, they are recognized using Support Vector Machine(SVM).

The previous studies ⁽³⁾⁽⁴⁾ have many examples using wet type sensors. Wet type sensors can obtain better data for precise recognition than dry ones. However, they need a high cost and are inconvenient to attach/detach. Therefore, in this paper we consider future practical use and propose a silent-voice recognition system of Japanese vowels by bipolar small dry type sensors.

2. Vowels recognition method

A method proposed in this paper consists of 4 parts, input part, data transform part, feature extraction part and train/recognition part (Fig. 1).

2.1 Input

In this paper, we use 3 electrodes to measure surface EMG signals. Electrodes which measures biological signals have roles as an interface of delivering electrifications between a living body and measurement devices. Biological signals are generally weak electric potentials which occur in a low frequency region. Therefore, measurement electrodes

Fig. 1. Process flow of proposed method.

must have a low electrode impedance and a stable electrode potential. In this subsection, we explain about a method of measuring surface EMG signals.

2.1.1 Location of sensors

We use 3 bipolar electrodes (Fig. 2) to measure surface EMG signals.

In the utterance of Japanese vowels, characteristic emerges in lips neighborhood conspicuously. We show states of mandible bone, lip and angulus oris at the time of the utterance of vowels in Table 1. From Table 1, we estimate suitable locations where sensors can read these utterance states. In this paper, 3 sensors are therefore used to measure surface EMG data at orbicularis oris muscle (1ch), zygomatic muscle (2ch) and depressor angle oris muscle (3ch) (Fig. 3). Orbicularis oris muscle has a role that lips make shrink, and then can generate EMG signals of all vowels. Zygomatic muscle's role is to pull angulus oris to outside of the upper side and then can generate EMG signals of "u" and "o". Depressor angle oris muscle's role is to pull angulus oris to lower and then can generate EMG signals of "i" and "e".

2.2 Data transform

EMG data is processed as described in this part. EMG data is partitioned for every utterance to make three frames. After that, a window function is applied to them. After windowing, FFT (Fast Fourier Transform) is applied to obtain frequency components (Fig. 4).

2.2.1 Decision of data range

After data input, the present system decides the range of EMG data used in the following processing. Voice data is generally divided into 3 sections, a Beginning, a Holding and a Closure sections (Fig. 5).

Fig. 2. Bipolar dry type sensors.

Table 1.State at 3 positions.

Vowels	Mandible bone	Lip	Angulus oris
a	Pull up	Open wide	
i	Hold	Open	Pull to outside
u	Hold	Shrink/Push	
e	Lower	Open	Pull to outside
o	Lower	Shrink/Push	

Fig. 3. Position of sensors.

A voice signal becomes steady state in the Holding section and the biggest characteristic appears in the Beginning section. In addition, my previous study showed the use of only the Beginning-section of EMG data cannot extract each characteristic.

Therefore, we use the Beginning-section's and Holding-section's data.

First, we watch amplitudes from a Beginning to a Holding-section of data to make absolute values of each channels' data and to combine them. After this, we seek a point where the maximum value is provided. Then, the proposed system assumes this point as a reference point

Fig. 4. Process flow of data transform part.

Fig. 5. Configuration of EMG data.

(Fig.6). After this, we extract 4,096 points (1,380 ms) data from 1,024 points before the reference point (Fig. 7).

2.2.2 Framing

After extracting data, the present system makes three frames. Data range used consist of 2,048 points and overlap range is 1,024 points (Fig. 8).

2.2.3 Windowing

After framing, a window function is applied to extracted data before frequency analysis. The window function is a hamming window (Fig. 9). This window function ($W(n)$) is expressed in (1), and has characteristics such as a narrower main lobe and lower side lobes than those of a Hanning window.

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (0 \leq n \leq N-1) \quad (1)$$

Fig. 6. How to decide a Reference point.

Fig. 7. Data range.

2.2.4 Frequency Analysis

In this paper, we use FFT for frequency analysis. Generally speaking, the frequency properties of EMG are distributed in the range of 0-500(Hz) ⁽⁵⁾. Therefore, we use this range's neighborhood data in this paper. We then obtain power spectra as data for feature extraction. Power spectra are shown in equation (2).

$$|F(t)| = \sqrt{\text{Re}(F(t))^2 + \text{Im}(F(t))^2} \quad (2)$$

2.3 Feature extraction

We use LDA (Linear Discriminant Analysis) for feature extraction and dimensional reduction. LDA is one of methods of supervised learning. Originally, it can train and recognize many kinds of data. However, the system in this paper uses LDA only for feature extraction.

Fig. 8. Framing.

Fig. 9. Hamming window.

2.4 Recognition

We use SVM (Support Vector Machine) for template

generation and recognition. SVM is one of methods of pattern recognition and has a characteristic that maximizes a margin between classes.

3. Experiment 1

3.1 Measurement of surface EMG signal

The subject is a man in twenties in this paper. In experiments, it is assumed that mounting sensors to face, measuring EMG, and demounting them are 1 trial. The subject utters 5 Japanese vowels 3 times (15 data/trial). It doesn't need actual utterance. The subject maintains the form of his mouth at pronunciation for 4 seconds and repeat each pronunciation 3 times (Table 2). The rest period is then arranged between utterances (Fig. 10).

3.2 Recognition

In these experiments, we measured EMG data 3 times. Then 2 times data (30 data) are used to make templates. The remaining 15 data are used for test. Accuracy evaluation is carried out by cross validation (Fig. 11).

In this paper, LDA reduces data dimension to 4. A kernel function of SVM is a Gaussian one. Parameters of SVM are $C=1$, $\gamma=0.25$. C means a cost parameter and γ is used to determine complexity of decision boundary.

3.3 Experimental result

We call 3 data 1st, 2nd and 3rd in the order that we measured. We then calculate recognition accuracy of each pattern and the average recognition accuracy of them.

As a result, we obtained 62.3% average recognition accuracy (Table 3).

3.4 Consideration

Recognition accuracy of the previous method is 33%. The proposed method improved recognition accuracy to 62.3%. Change of a feature extraction method and change of range at data extraction affected improvement of accuracy.

However, we cannot say that the accuracy is sufficiently high. The previous study using wet type sensors ⁽³⁾ obtained greater than 90% in accuracy. In comparison with this previous study, recognition accuracy of the proposed method is still bad.

We think the reasons for this phenomenon include the followings. First, in decision of data range, the proposed

Table 2. Measurement condition.

Objects to recognize	Japancsc 5 Vowels
The subject	A man in twenties
Number of determinations	3 times(45 data)
Measurement position	1. orbicularis oris muscle 2. zygomatic muscle 3. depressor angle oris muscle
Sensor	Bipolar dry type
Mcausing equipment	Personal – EMG ⁽⁷⁾ (Sampling rate : 3,000 Hz)

Fig. 10. Experimental flow.

Fig. 11. Dataset dividing of Experiment1.

Table 3. Result of the Experimentations.

Pattern (template/test)	recognition accuracy (%)
1 st , 2 nd / 3 rd	53.3%
1 st , 3 rd / 2 nd	66.7%
2 nd , 3 rd / 1 st	66.7%
AVERAGE	62.3%

system extracted from a Beginning to a Holding section's data. However, necessary features for recognition can be included in only the Holding section. Beginning section's data can include features to recognize only consonants. Therefore, the use of Beginning-section's data for vowels recognition is not suitable.

Second, the recognition results were maybe affected by intra-subject variations. In order to investigate the influence of intra-subject variations, we changed the value of C and gamma between 10^{-6} and 10^6 , parameters of SVM at the highest recognition accuracy were examined. We divide this range equally into 100 on a logarithmic scale and tried all

combinations. Table 4 means this experimental result. We can observe that the parameters significantly differ depending on the combination of the learning data. This difference of parameters shows experimental data have different features.

There is another problem concerning sensor types. Dry type sensors are hard to obtain good data like wet type ones. Fig. 12 shows raw EMG data. We can observe there are some noises in EMG data. In addition, noises caused by movement of sensors during measurement adversely affect EMG data. This can be the cause of noises.

In order to solve these problems, a new method of decreasing intra-subject variations and rejecting noise-like signals is necessary.

As next plans, we are going to increase the number of experimental data, change range of data extraction at the data transform and devise methods to decrease noise-like signals.

4. Experiment 2

After the experiment 1, we increased the number of subjects to 3 people. All subjects are men of twenties. Other conditions are the same as the experiment 1.

We changed overlap range at Data-transform part to 512 points, increased data dimension of experimental data from 3,096 to 5,160, and as a preprocessing of SVM, applied standardization that makes 0 mean and variance 1.

4.1 Datasets

This experiment use 135 data (45 data / subject). Then 2 subjects data (90 data) are used to make templates. The remaining subject's data (45 data) are used for test (Fig. 13). Accuracy evaluation is carried out by cross validation.

4.2 Experimental result

The subjects are called as A, B and C.

We confirm a difference of the subjects by changing parameters of SVM. We changed the values of C and gamma between 10^{-6} and 10^6 , and parameters of SVM at the highest recognition accuracy were examined. We divide this range equally into 100 on a logarithmic scale and tried all combinations. Table 5 shows this experimental result. As a result, the system obtained 48.9%, 40%, 55.6% best accuracy.

Table 4. Best parameters of SVM.

Pattern (Template/Test)	1 st , 2 nd / 3 rd	1 st , 3 rd / 2 nd	2 nd , 3 rd / 1 st
C	0.86975	305.38555	0.65793
Gamma	0.37649	0.00012	0.04037
Best accuracy	66.7%	73.3%	66.7%

Fig. 12. Raw EMG data.

Fig. 13. Datasets dividing of Experiment 2.

Table 5. Result of Experiment 2.

Pattern (Template/Test)	A, B / C	A,C / B	B,C / A
C	24.77076	75.646332	14.17474
Gamma	0.49770	0.00327	0.09326
Best accuracy	40%	55.6%	48.9%

4.3 Consideration

From this result, we can say this proposed method could not recognize EMG data by a subject using other subject's data. From parameters of SVM, it is shown the difference between subjects is large. That is to say, to increase accuracy we have to solve problems such as intra-subject variations and a difference between subjects. First, we must review data range of the Data-transform part. This system should use data ranges that do not show more individual characteristics and use/create data with less noise. Figs. 14 to 16 show raw EMG data at the time of pronunciation of "U" for each subject. Looking at these, we can see Fig. 14 has some noise-like signals and cannot distinguish the Holding-section, the Closure-section and the rest period. In addition, the EMG data in Fig. 16 is very small in amplitude and we cannot distinguish which range of data belongs to each section.

5. Conclusion

We proposed a system of vowels recognition using EMG obtained by bipolar dry type sensors. We used 3 dry type sensors and measured EMG signals in lips neighborhood. First, the reference point was determined to separate EMG data, and extracted 4,096 points data from 1,024 points before the reference point. Next, the Hamming window was applied to EMG data. After that, power spectra and phase spectra by FFT were obtained. LDA was applied to those data and reduced data dimension to four. After LDA, processed data were recognized by SVM with a Gaussian function. The subject is a man in twenties. We obtained 15 data (3 data/vowels) by one experiment. We performed it 3 times and obtained 45 data. In the experiment 1, the system used 45 data for template generation and recognition. Then 2 times data (30 data) are used to make templates. The remaining 15 data are used for test. Accuracy evaluation is carried out by cross validation. As a result, we obtained 62.3% average recognition accuracy. In the experiment 2, we increased the number of subjects to 3, and the system used 135 data for template generation and recognition. 90 data (2 subjects) are used for template generation and the remaining 45 data (1 subject's data) are used for test. In order to investigate a difference between subjects, we calculated best accuracy of each template pattern and compared a difference of parameters of SVM. From these results, it became clear we have to change range of data extraction at the data transform to decrease data difference between subjects and devise methods to decrease noise-like signals and so on.

References

- (1) Ryosuke Takabatake, Shin-ich Ito, Momoyo Ito and Minoru Fukumi: "Vowels recognition using EMG measured with dry type sensors", The International Conference on Electrical Engineering 2016, No.90049, 2016
- (2) Takeshi Saitoh, Mitsugu Hisagi and Ryosuke Konishi: "Japanese Phone Recognition using Lip Image Information," MVA2007 IAPR Conference on Machine Vision Applications, Tokyo, Japan, May 16-18, 2007
- (3) Manabe, H., A. Hiraiwa, and T. Sugimura: "Mime speech recognition: Recognition of 5 Japanese vowels without generating voice using electromyography.", *T. IEICE Japan* 88: 1909-1917, 2005
- (4) Fraiwan, Luay, et al: "Voiceless Arabic vowels recognition using facial EMG." *Medical & biological engineering & computing* 49.7: 811-818, 2011
- (5) Tomohiro Kizuka, Tadashi Masuda, Tohru Kiryu and Tsugutake Sadoyama, "Biomechanism Library Practical Usage of Surface Electromyogram," Tokyo Denki University Press, Japan, ISBN: 978-4-501-32510-7 C3047
- (6) Personal EMG, Oisaka Electronic Equipment Ltd. <http://www.oisaka.co.jp/P-EMG.html>

Fig. 14. EMG data of the subject A when utters vowel "U".

Fig. 15. EMG data of the subject B when utters vowel "U".

Fig. 16. EMG data of the subject C when utters vowel "U".