Japanese Vowel Sound Classification Using Fuzzy Inference System

Suwannee Phitakwinai[1], Hideyuki Sawada[2], Sansanee Auephanwiriyakul[1,3*], Nipon Theera-Umpon[3,4]

[1]Computer Engineering Department, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand, [2]Department of Intelligent Mechanical Systems Engineering, Faculty of Engineering, Kagawa University, Takamatsu, Japan, [3]Biomedical Engineering Center, Chiang Mai University, Chiang Mai, Thailand, [4]Department of Electrical Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

Abstract An automatic speech recognition system is one of the popular research problems. There are many research groups working in this field for different language including Japanese. Japanese vowel recognition is one of important parts in the Japanese speech recognition system. The vowel classification system with the Mamdani fuzzy inference system was developed in this research. We tested our system on the blind test data set collected from one male native Japanese speaker and four male non-native Japanese speakers. All subjects in the blind test data set were not the same subjects in the training data set. We found out that the classification rate from the training data set is 95.0 %. In the speaker-independent experiments, the classification rate from the native speaker is around 70.0 %, whereas that from the non-native speakers is around 80.5 %.

• **Key Words** : Vowel classification, Mamdani fuzzy inference system, Formant.

1. Introduction

An automatic speech recognition system is a phoneme identification process from the speech segment. This process relies heavily on a very good vowel sound classification system [1]. There are several research works on speech recognition system for different languages including Japanese. Therefore, Japanese vowel sound classification becomes one of the challenging research problems.

Since a very good vowel recognition system is an important part in speech recognition system, there are several research works on vowel recognition in different languages including English [2 - 5], English with Malaysian speakers [6], Arabic [7], Mandarin [8], Kazakh [9], Malay [10], Thai [11], etc. There are also some research works on the Japanese vowel sound (/a/, /i/, /u/, /e/, and /o/) recognition system [12 - 15]. Although these systems provide very good recognition rates, they may not work well with non-native speakers. Hence, in this paper, we develop the Japanese vowel classification system with the Mamdani fuzzy inference system. Our system is tested with a blind test data set collected from four non-native Japanese speakers and one native Japanese speaker.

This paper is organized as follows. The next section provides the descriptions of the vowel recognition system. The results are illustrated and discussed in section 3. Finally, the conclusion is drawn in section 4.

*교신저자 : Sansanee Auephanwiriyakul, (sansanee@ieee.org)
접수일 : 2014년 1월 7일, 수정일 2014년 2월 17일, 게재확정일 : 2014년 2월 22일

2. System Overview

The input to this system is Japanese vowel sound signals acquired at 8 kHz sampling rate. Each Japanese vowel sound signal is segmented into 512 samples. We preprocess each signal segment with the Hamming window [16] to improve its quality. The Hamming window is calculated as

$$w(n) = 0.54 - 0.46\cos\left(2\pi\frac{n}{N}\right) \tag{1}$$

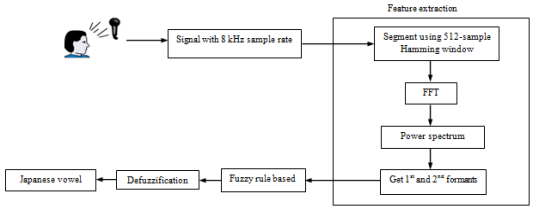
with N = 512. Then the resulting signal is transformed using the discrete Fourier transform [16], i.e.,

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-i2\pi kn}{N}}$$
,k = 1,2,⋯,N. (2)

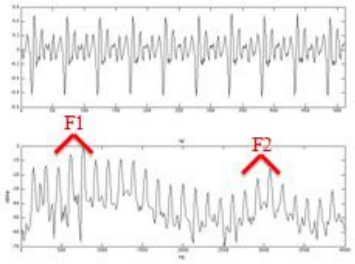
We calculated a power spectrum [16] of each signal by

$$P(k) = 20\log_{10}\left(\frac{|X(k)|}{\max_j |X(j)|}\right) . \tag{3}$$

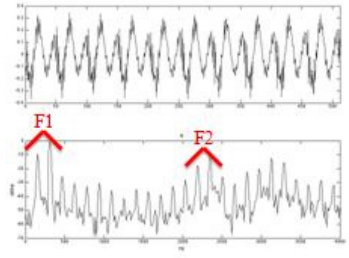
Then the Mamdani fuzzy inference system with the middle of maximum (mom) as a defuzzification method [17] is utilized to classify each Japanese vowel sound. The two input features of the system are the first (F1) and second (F2) formants [1, 16]. The summary of the process is shown in figure 1. Examples of each vowel sound and its first and second formants are shown in figure 2.



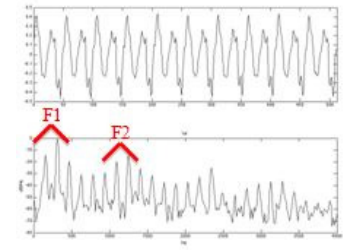
[Fig. 1] Summary of the recognition process



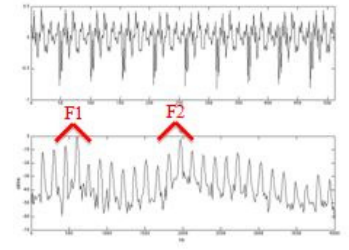
(a)



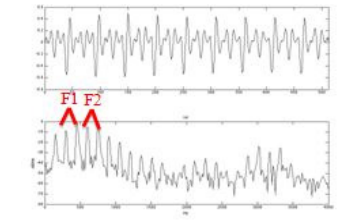
(b)



(c)



(d)

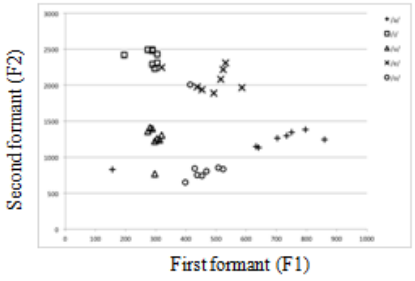


(e)

[Fig. 2] Japanese vowel signals and their corresponding first and second formants of vowel sound (a) /a/, (b) /i/ (c) /u/, (d) /e/, (e) /o/

3. Experimental Results

We collected a train data set from 4 male native Japanese speakers. Each of the 5 vowels was recorded 8 times and, hence, there were 40 samples in total for each subject. The first and second formants of all training samples are shown in figure 3.

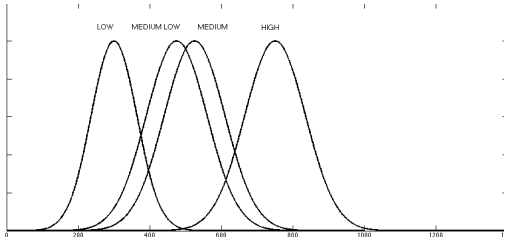


[Fig. 3] First and second formants of vowels in training set

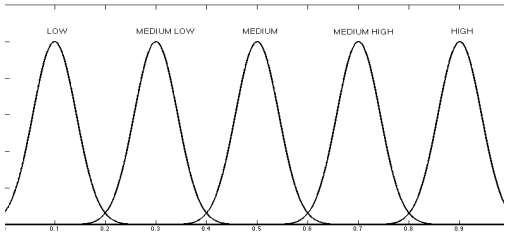
inference system as follows:

Rule 1: If (F1 is HIGH) and (F2 is MEDIUM) then (output is HIGH).
Rule 2: If (F1 is LOW) and (F2 is VERY HIGH) then (output is LOW).
Rule 3: If (F1 is LOW) and (F2 is MEDIUM) then (output is MEDIUM).
Rule 4: If (F1 is MEDIUM) and (F2 is HIGH) then (output is MEDIUM LOW).
Rule 5: If (F1 is MEDIUM LOW) and (F2 is LOW) then (output is MEDIUM HIGH).

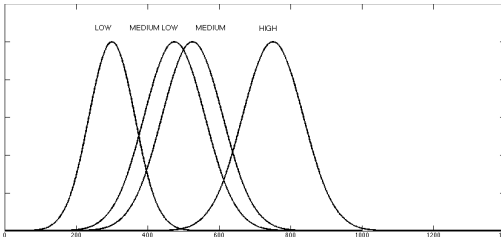
The membership function of F1, F2, and the output shown in figure 4 were manually constructed.



(a)



(b)



(c)

[Fig. 4] Membership functions of (a) F1, (b) F2, and (c) output

The defuzzified output from the mom method was used as an input to the manually selected classification rule bases shown as follows:

Rule 1: If an output ϑ [0.0, 0.2], then this vowel sound is /i/.
Rule 2: If an output ϑ [0.2, 0.4], then this vowel sound is /e/.
Rule 3: If an output ϑ [0.4, 0.6], then this vowel sound is /u/.
Rule 4: If an output ϑ [0.6, 0.8], then this vowel sound is /o/.
Rule 5: If an output ϑ [0.8, 1.0], then this vowel sound is /a/.

The confusion matrix of the training data set is shown in figure 5. We can see that the correct classification rate is 95.0 %. Two samples of vowel /o/ are misclassified as /e/. The reason is that the first and second formants of those samples are close to those of vowel /e/. For example, the first and second formants of one of the misclassified /o/ are around 625 and 2679.7 Hz. Hence, the defuzzified output from the

Mamdani is around 0.3 and that will assign this sample to /e/ group. For another misclassified /o/, the system gives the defuzzified output approximately 0.3, as well

		Algorithm's output				
		/a/	/i/	/u/	/e/	/o/
Actual vowel	/a/	8				
	/i/		8			
	/u/			8		
	/e/				8	
	/o/				2	6

[Fig. 5] Confusion matrix from train data set

We tested our system with a male native Japanese speaker and 4 male non-native Japanese speakers. These 5 subjects are not the same subjects in the training data set. We collected each vowel sound 10 times from a native speaker, hence, there were 50 samples in total in the first blind test data set. For the second blind test data set, we also asked each non-native Japanese speaker to speak each vowel 10 times, hence, there were 50 samples from each speaker. Therefore, there were 200 samples in total. The confusion matrix of the native and non-native Japanese speakers is shown in figures 6 and 7. In the speaker-independent experiments, the correct classification rate on the native Japanese speaker is 70.0 %, while the total correct classification rate of the non-native speakers is 80.5 %. Again, the misclassified vowels are caused by the same reason as mentioned before. For example, the first and second formants of vowel /o/ from the native Japanese speaker shown in figure 8 are close to those of vowels /u/ and /a/ in the training data set. The defuzzified outputs of this misclassified /o/ are approximately 0.9 (in /a/ group) and around 0.5 (in /u/ group). The first and second formants of vowel /e/ from the first non-native speaker shown in figure 9 are around those of vowel /i/ in the training data set. The defuzzified output is approximately 0.1 that will result in the assignment of this sample to /i/ group.

		Algorithm's output				
		/a/	/i/	/u/	/e/	/o/
Actual vowel	/a/	8				2
	/i/		8	1	1	
	/u/			10		
	/e/	2		1	7	
	/o/	2		6		2

[Fig. 6] Confusion matrix of blind test data set from a native Japanese speaker

		Algorithm's output				
		/a/	/i/	/u/	/e/	/o/
Actual vowel	/a/	10				
	/i/		10			
	/u/		5	1	4	
	/e/		10			
	/o/					10

(a)

		Algorithm's output				
		/a/	/i/	/u/	/e/	/o/
Actual vowel	/a/	9				1
	/i/		9	1		
	/u/			8	2	
	/e/				10	
	/o/				1	9

(b)

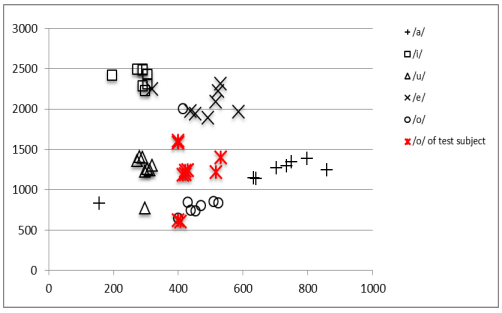
		Algorithm's output				
		/a/	/i/	/u/	/e/	/o/
Actual vowel	/a/	10				
	/i/		7	2		1
	/u/			10		
	/e/				10	
	/o/					10

(c)

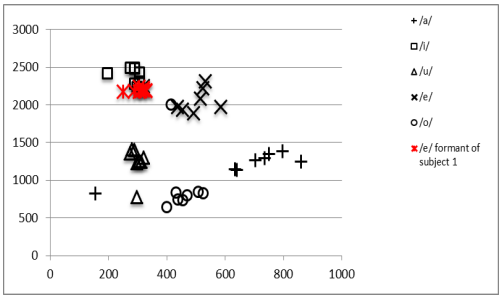
		Algorithm's output				
		/a/	/i/	/u/	/e/	/o/
Actual vowel	/a/	6			2	2
	/i/		7		3	
	/u/		1	5	3	1
	/e/				10	
	/o/					10

(d)

[Fig. 7] Confusion matrices of blind test data set from non- native Japanese speakers (a) 1, (b) 2, (c) 3, and (d) 4



[Fig. 8] First and second formants of vowel /o/ of the blind native speaker among formants of all vowels in training data set



[Fig. 9] First and second formants of vowel /e/ of the first blind non-native speaker among formants of all vowels in training data set

4. Conclusion

There are several research works on automatic speech recognition system. The Japanese speech recognition system is one of the popular research topics. One of the important parts in the research topic is Japanese vowel recognition. In this paper, we developed a vowel recognition system using the Mamdani fuzzy inference system. We found that the recognition rate of the training data set was approximately 95.0 %. The correct classification rate of the blind test data collected from a native Japanese speaker was around 70.0 %, whereas those of 4 non-native Japanese speakers was approximately 80.5 %. Hence, the results demonstrated that this system worked well in the speaker-independent experiments, even when the speakers were not native speaker. We will integrate this system into the complete speech recognition system in the future work.

Acknowledgments

This work was supported by the Japan Student Services Organization (JASSO)

REFERENCES

[1] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, New Jersey, 1993.

[2] D. G. Kimber, M. A. Bush, and G. N. Tajchman, "Speaker-independent vowel classification using Hidden Markov Models and LVQ2," *International Conference on Acoustics, Speech, and Signal Processing*, pp. 497-500, 1990.

[3] T. Harczos, G. Szepannek, A. Katai, and F. Klefenz, "An auditory model based vowel classification," *IEEE Biomedical Circuits and Systems Conference*, pp. 69-72, 2006.

[4] A. Sadeghian, H. R. Dajani, and A. D. C. Chan, "Classification of English vowels using speech evoked potentials," *33rd Annual International Conference of the IEEE EMBS*, pp. 5000-5003, 2011.

[5] J. Hillenbrand and R. T. Gayvert, "Vowel classification based on fundamental frequency and formant frequencies," *Journal of Speech and Hearing Research*, Vol. 36, pp. 694-700, 1993.

[6] P. M. Paulraj, S. B. Yaacob, A. Nazri, and S. Kumar, "Classification of vowel sounds using MFCC and feed forward neural network," *5th International Colloquium on Signal Processing & Its Application*, pp. 59-62, 2009.

[7] K. Daqrouq, K. Y. Al Azzawi, "Arabic vowels recognition based on wavelet average framing linear prediction coding and neural network," *Speech Communication*, Vol. 55, pp. 641-652, 2013.

[8] C.-T. Hsieh, E. Lai, and Y.-C. Wang, "Distributed

fuzzy rules for preprocessing of speech segmentation with genetic algorithm," *IEEE International Conference on Fuzzy Systems*, pp. 427-431, 1997.

[9] Z. Yessenbayev, M. Karabalayeva, and A. Sharipbayev, "Formant analysis and mathematical model of Kazakh vowels," *14th International Conference on Computer Modelling and Simulation*, pp. 427-431, 2012.

[10] A. Zourmand, and T. H. Nong, "Vowel classification of children's speech using fundamental formant frequencies," *4th International Conference on Computational Intelligence, Modelling and Simulation*, pp. 282-287, 2012.

[11] N. Theera-Umpon, S. Chansareewittaya, and S. Auephanwiriyakul, "Phoneme and Tonal Accent Recognition for Thai Speech," *Expert Systems With Applications*, Vol. 38, No. 10, pp. 13254-13259, 2011.

[12] S. Karungaru, T. Kamei, M. Fujiwara, and N. Akamatsu, "Vowel recognition using Akamatsu integral and Differential Transforms," *International Journal of Advanced Intelligence*, Vol. 1, No. 1, pp. 125-140, 2009.

[13] M. Hisagi, T. Saitoh, and R. Konishi, "Analysis of efficient feature for Japanese vowel recognition," *International Symposium of Intelligent Signal Processing and Communication Systems*, 2006.

[14] T. Murakami, K. Maruyama, N. Minematsu, and K. Hirose, "Japanese vowel recognition using external structure of speech," *IEEE Workshop on Automatic Speech Recognition and understanding*, pp. 33-36, 2005.

[15] T. Murakami, K. Maruyama, N. Minematsu, and K. Hirose, "Japanese vowel recognition based on structural representation of speech," *Annual Conference of International Speech Communication Association*, pp. 1261-1264, 2005.

[16] A. V. Oppenheim and R. W. Schafer, *Discrete-time Signal Processing*, Prentice Hall, New Jersey, 1989.

[17] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

저자소개

Hideyuki Sawada

- 1990. March : Waseda University, Japan, Applied Physics, (B. Eng.)
- 1992. March: Waseda University, Japan, Applied Physics, (M.S.)
- 1999. February:Waseda University, Japan, Dr. Eng.
- 1999. Apr.~2010. March : Associate Professor of Dept. of Intelligent Mechanical Systems Engineering, Kagawa University, Japan
- 2010 Apr.~Present : Professor of Dept. of Intelligent Mechanical Systems Engineering, Kagawa University, Japan
- E-Mail : sawada@eng.kagawa-u.ac.jp

<Research Interest> :
Sound and image processing, Neural networks, Robotics, Human interfaces and Assistive technologies

Sansanee Auephanwiriyakul

- 1993. Feb. : Chiang Mai University, Thailand, Electrical Engineering, (B. Eng. (Hons.))
- 1996. July :University of Missouri-Columbia, U.S.A., Electrical and Computer Engineering (M.S.)
- 2000. December: University of Missouri-Columbia, U.S.A., Computer Engineering and Computer Science (Ph.D.)
- 2001. January ~ August : Post-Doctoral Fellow, University of Missouri-Columbia, U.S.A., Computer Engineering and Computer Science (Ph.D.)
- 2002. January ~ Present : Assistant Professor of Dept. of Computer Engineering, Chiang Mai University, Thailand
- E-Mail : sansanee@ieee.org

<Research Interest> :
Pattern Recognition, Digital Image Processing, Computer Vision, Neural Networks, Fuzzy Logic, Fuzzy set and system, Type-2 fuzzy set, Medical Signal and Image Processing

Nipon Theera-Umpon

- 1993. Feb. : Chiang Mai University, Thailand, Electrical Engineering, (B. Eng. (Hons.))
- 1996. May : University of Southern California, U.S.A., Electrical Engineering (M.S.)
- 2000. May : University of Missouri-Columbia, U.S.A., Electrical Engineering (Ph.D.)
- 1993. Apr. ~ Present : Associate Professor of Dept. of Electrical Engineering, Chiang Mai University, Thailand
- E-Mail : nipon@ieee.org

<Research Interest> :
Pattern Recognition, Digital Signal Processing, Digital Image Processing, Neural Networks, Fuzzy Logic, Medical Signal and Image Processing