

# The Features of Vowel /E/ Pronounced by Chinese Learners

Yasukazu Kanamori

Graduate School of Information Science and Technology, Aichi Prefectural University, Nagakute, Japan

Email: kanamori@ist.aichi-pu.ac.jp

Guoxing Fang

Toyota Communication Systems Co., Ltd, Nagoya, Japan

Email: k-ho@toyota-cs.com

**Abstract**—In this paper, we investigate the features of the syllables including vowel /e/, /u/ and /ü/ which are pronounced by Chinese learners. How the pronunciation error of vowel /e/ being influenced by position of two-syllable word, tone as well as consonant is mainly discussed. As results, firstly it is understood that the position of two-syllable word does not affect the pronunciation of vowel /e/, and secondly, for tone, the pronunciation error rate is highest when the syllable is the third of tone. In order to objectively judge the pronunciation state, we used the value of the first formant frequency in the beginning, middle and ending part of vowel /e/. The proposed method is confirmed by obtaining 80.1% correct rate when comparing the pronunciations with perception result between native Chinese students and Chinese learners of Japanese students using 54 Chinese words.

**Index Terms**—pronunciation, vowel /e/, Japanese student, verification, first formant frequency F1

## I. INTRODUCTION

Recently, with the development of the economics and internationalization of China, Chinese learners have increased rapidly. However, pronunciation of Chinese education faces with a lot of difficulties because of complex phonetics [1]. Pronunciation errors are often made by foreign language learners. Especially when the target language contains some phonemes that are not found in learners' native language, learners will use these phonemes with ones existing in their native language. Some katakana-English dictionaries are good example for Japanese. Hence, there are many researchers have directly investigated the characteristics of pronunciation of Chinese [2]-[4] and approved some information processing techniques [5], [6]. The single vowel analysis and Computer Assisted Language Learning (CALL) [7] have been actively researched. However most of these researches are about pronunciation of consonant and prosody. There is very few research of two-syllable word of vowel about Chinese learners of Japanese students. Automatic detection of these errors is one of essential and requisite technique in CALL systems [1].

Chinese characters are composed with vowels, consonants and tones. A single-syllable word is the basic of the Chinese. It is composed with a consonant and a vowel. There are basically six types of single-vowel including /a/, /o/, /e/, /i/, /u/ and /ü/. There are also diphthongs and trip thongs that composed from single-vowel. Therefore, if learners want to pronounce Chinese correctly, it is have to, first of all, pronounce single-vowel correctly. Otherwise, there are four tones as Chinese normal tone and a neutral tone which is pronounced shortly and lightly, usually called as zero-tone.

In this paper, we investigated the features of the syllables including vowel /e/, /u/ and /ü/ which are pronounced by Chinese learners. Single-vowels of /e/, /u/ and /ü/ are generally considered to be difficult for Chinese learners of Japanese students to pronounce. We investigated the pronunciation of /e/, /u/ and /ü/ which are included in two-syllable word. It is understood that /e/ is the most mistaken pronounced single-vowel among /e/, /u/ and /ü/. Therefore we study the single-vowel /e/ in detail. How the pronunciation error of vowel /e/ being influenced by position of two-syllable word, tone as well as consonant is mainly discussed [8].

## II. CHINESE AUDIO MATERIAL

In order to investigate the state of the pronunciation of Chinese learners, we recorded 126 two-syllable words. Nine Japanese students who have learned Chinese for 2 and 3 years are speakers. In this paper, "2" and "3" means the Japanese students who learned Chinese for 2 and 3 years, respectively. But in this paper we don't consider the difference in the level of year. For comparison with the Chinese learners of Japanese students, we also recorded the same words from 5 native Chinese speakers. There are 54 words contain the vowel /e/, 42 words contain the vowel /ü/, and 13 words contain the vowel /u/. The words being chosen to be recorded are in account of the balance in front and rear positions of the vowel. Table I shows the details of configuration of the word.

Audio data obtained from Chinese learners of Japanese students and native Chinese were recorded using a digital audio recorder (Roland R-09HR) under the following

conditions: sampling frequency = 48kHz, number of quantization bits = 16 bit. Five native Chinese speakers were asked to hear the data and give their evaluation results. We used three-level evaluation to evaluate the pronunciation here. The accuracy of pronunciation is showed in the Table I. From Table I, we can see that the error rate of vowel /e/ is near 50% and is higher than the others, so it is necessary to analyze vowel /e/ in more detail.

TABLE I. CONFIGURATION OF THE WORD

|   | Number of syllable | Number of words | Position of syllable |    | Error rate [%] |
|---|--------------------|-----------------|----------------------|----|----------------|
| e | 55                 | 54              | first                | 27 | 49             |
|   |                    |                 | second               | 28 | 55             |
| ü | 42                 | 42              | first                | 20 | 18             |
|   |                    |                 | second               | 22 | 20             |
| u | 13                 | 13              | first                | 7  | 13             |
|   |                    |                 | second               | 6  | 7              |

### III. ANALYSIS OF VOWEL /E/ PRONOUNCED BY CHINESE LEARNERS OF JAPANESE STUDENTS

Five native Chinese heard the audio data pronounced by the Chinese learners. Evaluation was done by taking three levels: wrong, ambiguous and correct pronunciations for each vowel in two-syllable word with two vowels. Fig. 1 shows the pronunciation error rate of /e/ of Chinese learners. From Fig. 1, we can see that the error rate of 6 speakers is greater than 50% in total 9 speakers which implies that it is difficult to pronounce vowel /e/ correctly. The symbols with the initial 3 of 3A to 3E is indicated 5 speakers of 3 years undergraduate students, and the symbols with the initial 2 of 2A to 2D is indicated 2 years students, respectively in Fig. 1. The difference between the grades is not so clearly.

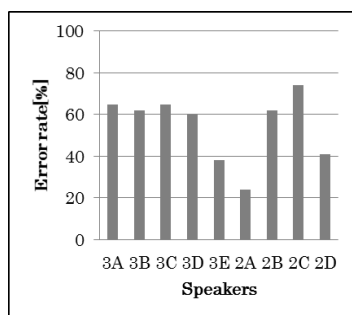


Figure 1. Error rate of vowel /e/

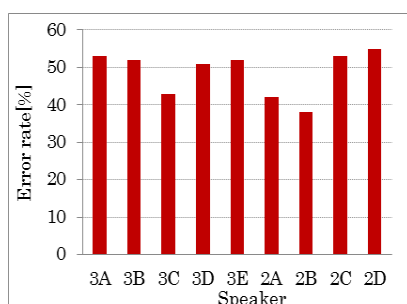


Figure 2. Error rate of first syllable with /e/ to whole word

Fig. 2 shows the error rate of vowel /e/ that is located at the first syllable of a two-syllable word. From the Fig. 2, we found that there is small influence on error rate when /e/ is in the first syllable position to six students. The ratio is a little higher for other three students in the second syllable.

### IV. RELATIONSHIP BETWEEN TONE AND PRONUNCIATION OF VOWEL /E/

Fig. 3 shows the relationship between tone and error rate of pronunciation of /e/. From Fig. 3, we find out that the error rate is the biggest for the third-tone, and that the error rate, which is lower than 15%, is the smallest for the zero-tone. This means that it is relatively easier to pronounce the zero-tone than the third tone.

Fig. 4 shows how the error rate of /e/ being influenced by tones for individual speaker. It is seen that error rate is a variable from person to person. However, the error rates of most Chinese learners of Japanese students are in a lower level for the first-tone and the second-tone, then the rates rise at the third-tone. From Fig. 3 and Fig. 4, the error rate of the third-tone is 62.7% which is the highest rate among the discussed tones. This result reflects that the third-tone itself is the most difficult one to pronounce in the tones of Chinese. In addition to combine the /e/, it was considered more difficult to pronounce. The lowest error rate of /e/ is zero-tone, and its average error rate is only 14.7%. The reason is that there is feature by syllable. For example, such as special syllable “的 [de]” and neutral tone is often to be pronounced when followed by the same syllable in the word. It is almost lower than the one of other tones for each speaker.

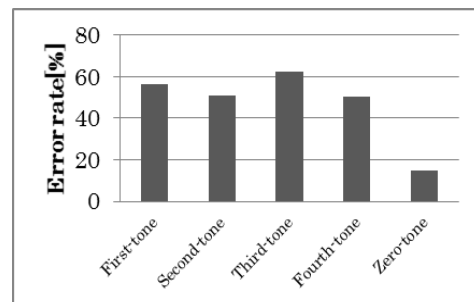


Figure 3. Relationship between error rate and tone of /e/

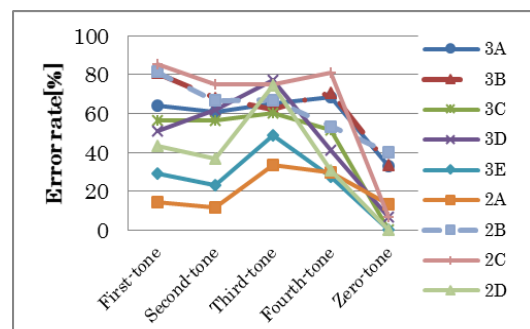


Figure 4. The influence of tones for individual speaker

The Error Rate (ER) of /e/ for each proceeding consonant is shown in Table II. It is the highest when consonant /r/ was combined with vowel /e/ while the error rate is the lowest

when consonant of /zh/ was combined with vowel /e/. This can be explained by considering that Japanese students do not used to pronounce the combination of /e/ and /r/ because there is almost no this combination in Japanese language, while Japanese students are familiar with zero-tone pronunciation and furthermore it is easy for Japanese students to pronounce when the consonant /zh/ is located at the second-syllable of a word. Fig. 5 shows the error rate's details of Table II in the figure.

TABLE II. ERROR RATE (ER) OF /E/ FOR EACH PROCEEDING CONSONANT

| Cons. | Numb | ER (%) | First syllable |        | Second syllab |        |
|-------|------|--------|----------------|--------|---------------|--------|
|       |      |        | Numb           | ER (%) | Numb          | ER (%) |
| c     | 4    | 42.9   | 2              | 44     | 2             | 41.8   |
| g     | 18   | 50.4   | 9              | 46.8   | 9             | 54.1   |
| k     | 13   | 55.5   | 4              | 44.4   | 8             | 62.4   |
| h     | 5    | 54.4   | 4              | 46.1   | 1             | 62.7   |
| l     | 1    | 33.3   | 1              | 33.3   | 0             | 0      |
| r     | 8    | 79.3   | 5              | 73.3   | 3             | 89.3   |
| zh    | 3    | 5.6    | 0              | 0      | 3             | 5.6    |
| ch    | 2    | 54     | 1              | 54.7   | 1             | 53.3   |
| sh    | 1    | 54.2   | 1              | 54.2   | 0             | 0      |

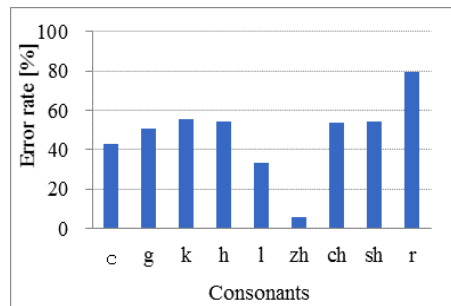


Figure 5. Error rate of vowel /e/ for each consonant

The consonants can be divided into two parts i.e. first and second syllable in this study. The relationship of error rate between these consonants and vowel /e/ has been investigated and the results are shown in Fig. 6. From the figure, we find out that the second consonant presents a higher average error rate than the first one [4].

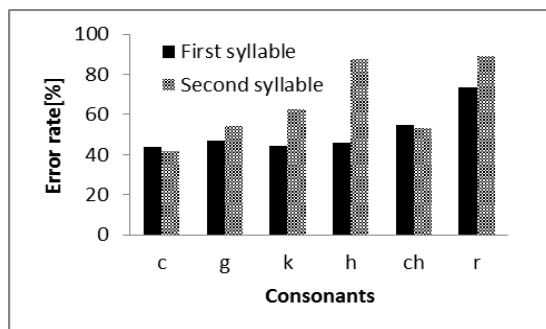


Figure 6. Comparison of the first and the second consonants

## V. DISTINCTION OF PRONUNCIATION

In order to distinguish automatically whether vowel /e/ is pronounced correctly or not to use computer as a tool, we have investigated the difference between the

pronunciation of native Chinese speakers and Japanese students. The vowel /e/ section is characterized by first formant frequency (F1). In order to investigate the variation in the vowel /e/, we quantify the data as follows:

1. Calculation of the overall F1 of vowel /e/ interval,
2. Dividing vowel /e/ into three sections (beginning, middle, ending) using each 4 frames,
3. Calculation of the average of F1 of each interval.

Fig. 7 shows an example of extraction of F1 that pronounced by Japanese student. The analysis condition of formant is shown in Table III.

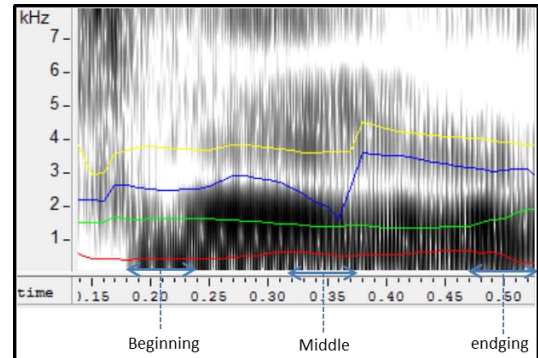


Figure 7. First formant of /e/ vowel

TABLE III. ANALYSIS CONDITION OF FORMANT FREQUENCY

|                    |            |
|--------------------|------------|
| Sampling frequency | 16kHz      |
| Frame length       | 410 Points |
| Shift interval     | 205 Points |
| Window type        | Hamming    |
| LPC order          | 24         |

## VI. VARIATION IN THE F1 FOR EACH TONE

Fig. 8 shows the analysis results of the first-tone which are pronounced by 2 native Chinese speakers and 2 Japanese students.

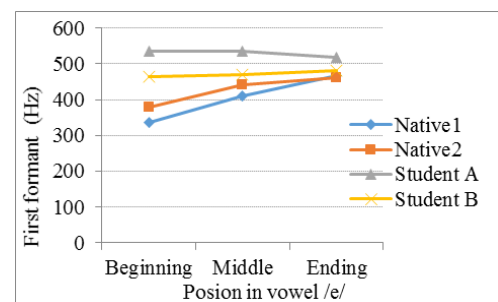


Figure 8. Comparison between native Chinese speakers and Japanese students for the first-tone of /e/

Vowel /e/ pronounced by two Japanese students A and B are evaluated as “correct” by the perception. First formant frequencies of two native Chinese speakers show a rising from beginning to ending for the first-tone of /e/. On the other hand, formants of Japanese students keep almost unchanged. Fig. 9 shows the analysis results of the second-tone. Japanese student A is evaluated to be ambiguous level, and student B is evaluated to be mistaken level. For student A, the value does not change between the beginning and middle and goes up from

middle to ending. For student B, the value goes down from middle to ending. Values of native Chinese speakers keep rising from beginning through middle to ending.

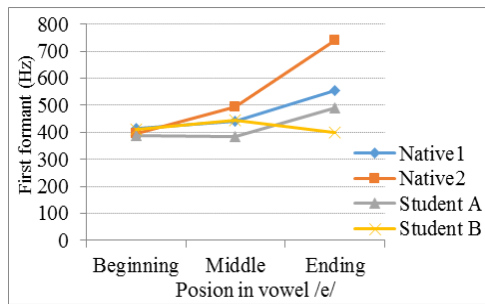


Figure 9. Comparison between native Chinese speakers and Japanese students in the second-tone of /e/

TABLE IV. DISCRIMINATION RULE

| Discrimination rule    | Decision result   |
|------------------------|-------------------|
| $R1 > 0 \wedge R2 > 0$ | Correct pronounce |
| $R1 < 0$ or $R2 < 0$   | Mistake pronounce |

We use Table IV to distinct the pronunciation of vowel /e/. In here, R1 means the first formant frequency F1 value after subtracting the beginning from the middle, and R2 means the F1 value of subtracting the middle from the ending.

#### VII. VERIFICATION OF THE PROPOSED METHOD

Fig. 10 shows the analysis results of 54-word pronunciations obtained from 4 Japanese students and 2 native Chinese speakers.

In Fig. 10, data of native Chinese speakers and correctly pronounced by Chinese learners lay mostly in the upper right corner while the data of non-correctly pronounced by Japanese students distribute in the lower left corner. The accuracy rate of 80.1% is obtained.

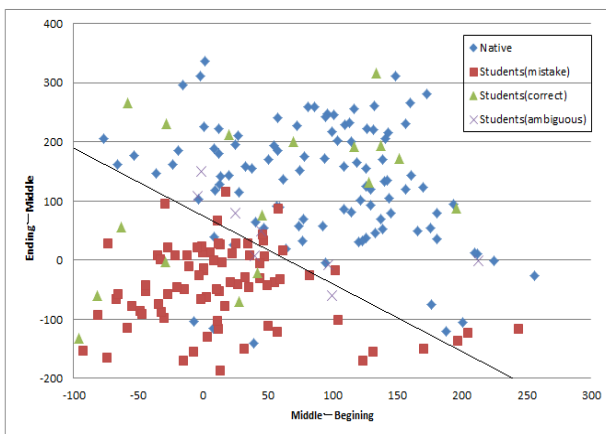


Figure 10. Distributions of R1 and R2

#### VIII. CONCLUSION

We investigated the characteristics of the syllables including vowel /e/, /u/, /i/ of Chinese pronunciation. An

effort was put on investigation of vowel /e/ since it is the most difficult to be pronounced among the vowels for Chinese learners of Japanese students. From the investigations, we found out that the error rate of /e/ is higher than 50%. By discussing how the pronunciation error of vowel /e/ being influenced by position of two-syllable word, tone as well as consonant, we found out the following conclusions: 1) relationship between accuracy of pronunciation and position of word does not really matter, 2) the error rate is the lowest at zero-tone, and the error rate is highest at the third-tone, 3) the error rate is the lowest for combining consonant /zh/ with vowel /e/, and the error rate is the highest when combining consonant /r/ with vowel /e/.

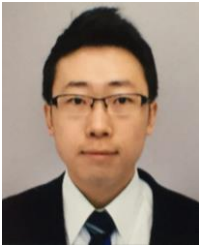
The proposed method of use R1 and R2 to distinct the state of the utterance obtained the accuracy rate of 80.1%. Making a fully automatic self-learning system is our next subject in the near future.

#### REFERENCES

- [1] Z. Zhang and S. Makino, "Chinese vowel recognition of using formant," *Acoustical Society of Japan*, vol. 47, no. 4, 1991.
- [2] Y. Kanamori, "The characteristics of Chinese vowel 'an' and 'ang' for Japanese learner," in *Proc. 18th International Congress on Acoustics*, 2004.
- [3] Y. Kanamori and T. Tokoro, "The feature extraction and discrimination of Chinese aspirated and un-aspirated affrications," *GESTS International Transaction on Computer Science and Engineering*, vol. 8, no. 1, pp. 105-111, May 2005.
- [4] S. C. Tseng, K. Kuei, and P. C. Tsou, "Acoustic characteristics of vowels and plosives/affricates of Mandarin-speaking hearing-impaired children," *Clinical Linguistics & Phonetics*, vol. 25, no. 9, pp. 784-803, 2011.
- [5] M. Eskenazi, "An overview of spoken language technology for education," *Speech Communication*, vol. 51, pp. 832-844, 2009.
- [6] T. Zhao, T. Zhao, A. Hoshino, M. Suzuki, N. Minematsu, and K. Hirose, "Automatic Chinese pronunciation error detection using SVM trained with structural features," in *Proc. Spoken Language Technology Workshop*, 2012, pp. 473-476.
- [7] T. Takagi, A. Hattori, and M. Komiya, "A Chinese language learning system with visualization and speech correction for prosody," *IEICE Trans.*, vol. J88-D-I, no. 2, pp. 478-487, 2005.
- [8] X. Yang and F. Gao, "An acoustics experiment of vowel duration in Chinese," *Bulletin of Hokkaido Bunkyo University*, vol. 29, pp. 65-79, 2005.
- [9] S. Hiki and K. Imaizumi, "A CAI system for self-teaching Chinese tones based on their acoustical properties," *IEICE Technical Report*, Sp2005-41, 2005.



**Yasukazu Kanamori** received the B.S. degree in electric engineering from Nanjing University of Science and Technology, Nanjing, China in 1982. Then he received the M.D. and Ph.D. degree in Graduate School of Engineering from Utsunomiya University, Utsunomiya, Japan in 1990 and 1996. From 1990 to 1993 and 1996 to 2000, he was an Assistant Professor of Utsunomiya University and Nara Institute of Science and Technology, respectively. From October 2000 to March 2002, he was a Visiting Researcher at Advanced Telecommunications Research Institute International, Spoken Language Translation Research Laboratories (ATR-SLT). He is currently an Associate Professor at Graduate school of Information and Science Technology, Aichi Prefectural University. His research interests include speech and audio signal processing, foreign language learning assistance, and dialect features analysis. Dr. Kanamori is a member of IEICE and ASJ of Japan.



**Guoxing Fang** received the B.S. degree in College of Engineering from Chubu University, Nagoya, Japan in 2009. Then he received the M.S. degree in Graduate school of Information and Science Technology, Aichi Prefectural University, Nagakute, Japan in 2012. He is currently an engineer at Toyota communication systems Co., Ltd. He was interested in foreign language learning when he was the student.