ACOUSTIC FEATURES OF JAPANESE WORDS SPOKEN BY JAPANESE NATIVES AND NON-NATIVES

Kimiko Yamakawa[1], Shigeaki Amano[2], Mariko Kondo[3]

[1]Shokei University, [2]Aichi Shukutoku University, [3]Waseda University
jin@shokei-gakuen.ac.jp, psy@asu.aasa.ac.jp, mkondo@waseda.jp

ABSTRACT

To investigate acoustic feature differences between Japanese native- and non-native speakers in spoken Japanese, 29 Japanese words were digitally recorded by 10 each of Japanese, French, Korean, Thai, Taiwanese, and Vietnamese speakers. The results revealed that acoustic features in each language have different tendencies in relative duration, intensity, and fundamental frequency. For example, compared to Japanese natives, Thai and Vietnamese speakers had the tendency of longer relative duration for first and last vowels than Japanese speakers, French speakers had lower relative intensity for all vowels, and all non-native speakers had higher relative fundamental frequency for first vowels but lower relative fundamental frequency for last vowels. It is suggested that the characteristics of each language of non-native speakers cause respective effects on the production of Japanese speech.

Keywords: Japanese, non-native speakers, duration, intensity, fundamental frequency

1. INTRODUCTION

Non-native speakers of Japanese sometimes mispronounce Japanese phonemes [5 , 8]. This mispronunciation is one cause of the unnaturalness of their Japanese speech. However, even if non-native speakers make no mispronunciations, their speech is often perceived as unnatural by Japanese native speakers [9]. It can be said that Japanese native speakers are sensitive to the naturalness of Japanese speech.

Previous research has suggested that the naturalness of Japanese speech relates to various acoustic features such as the duration, intensity, and fundamental frequency of speech segments [1, 2, 3, 4, 7]. For example, Sato [7] showed that the most effective factor for naturalness is a fundamental frequency in Japanese speech produced by Chinese and Korean speakers. Amano et al. [1] reported that Japanese speech by Chinese and Japanese speakers can be discriminated with standard deviation and the coefficient of variation in speech segment duration. Their results suggest that the difference in these durational features causes unnaturalness in Japanese speech by Chinese speakers.

However, each of the previous studies have only used a few languages of non-native speakers, and have analyzed different acoustic features using different methods. No unified analysis has been conducted for acoustic features across multiple languages of non-native speakers. In this context, the current study systematically examines acoustic features such as duration, intensity, and fundamental frequency in Japanese speech produced by non-native speakers such as French, Korean, Thai, Taiwanese, and Vietnamese to reveal their difference from Japanese native speakers. The current study focused on vowels not consonants in Japanese speech, because fundamental frequency cannot be obtained when a consonant is voiceless.

2. SPEECH RECORDING

2.1. Word materials

The word materials comprised 29 Japanese words that were two or three mora long and did not contain special moras such as lengthened vowels, geminate stops, or moraic nasals. The words all had flat type accents.

2.2. Speakers

Ten Japanese native speakers and 50 non-native speakers (10 each of French, Korean, Thai, Taiwanese, and Vietnamese) participated in the recording. The non-native speakers could at least read Hiragana, a moraic alphabet for the Japanese language. The number of male and female speakers and their ages are shown in Table 1. They were paid for their participation in the recording.

2.3 Equipment

Recording equipment included a microphone (ECM-999, SONY), an A/D converter (UA-25EX, Roland), and a personal computer (Dynabook SSRX2/T9L, TOSHIBA) with an extended display (SK-DTV133JW2, SK-NET, or On-Lap 1302/J, GECHIC) and a flexible keyboard (CB-KB85U02-BK, CLEVERY). A linear PCM recorder (DR-1,

Table 1 Characteristics of speakers participating in the recording

Native Language	Number of speakers			Age (year)			
	Male	Female	Total	Mean	Min.	Max.	SD
Japanese	4	6	10	24.1	20	20	3.5
French	4	4	10	21.9	20	25	1.6
Korean	3	7	10	20.7	19	23	1.5
Thai	0	10	10	21.6	20	23	1.0
Taiwanese	4	6	10	22.7	21	26	1.4
Vietnamese	8	2	10	21.8	21	23	0.7
Total	23	35	60	22.1	19	30	2.1

TASCAM) connected to a microphone (ECM-999, SONY) provided a simultaneous backup recording.

2.3. Procedure

Recording was conducted in a quiet room. The word material embedded in a carrier sentence, /korewa __ dato omoi masu/ ("I suppose that this is __") was presented on a computer display in Hiragana. A speaker pronounced the presented sentence at a normal speaking rate. The pronunciation was digitally recorded with 16-bit quantization and 48-kHz sampling frequency. The recording was then stored as a digital audio file on a computer. Each speaker pronounced each word once. The word order was randomized for each speaker.

The recording venues were Bordeaux in France for French speakers, Sangmyung in Korea for Korean speakers, Taipei in Taiwan for Taiwanese speakers, Bangkok in Thailand for Thai speakers, and Hanoi and Ho Chi Minh City in Vietnam for Vietnamese speakers. Native speakers of Japanese were recorded in Tokyo in Japan.

3. ANALYSIS

3.1. Procedure

Expert labelers added phoneme labels to the recorded words. Using these phoneme labels, the words and their vowels were extracted and analyzed.

The duration of vowels and words was calculated by subtracting their start time from their end time. The relative duration of vowels (*rD*) was calculated by dividing the vowel duration (*Dv*) by the word duration (*Dw*) (Eq. 1).

$$(1) \qquad rD_n = \frac{Dv_n}{Dw}$$

where *n* is the vowel position in the word.

The intensity of vowels and words was calculated using a rectangular window of 6-ms width with 1-ms shift and expressed in dB with 1 in the 16-bit signed integer as the reference level. The relative intensity of a vowel (*rI*) was calculated by subtracting the word's averaged intensity (*mIv*) from the vowel's averaged intensity (*mIw*) (Eq.2).

$$(2) \qquad rI_n = mIv_n - mIw$$

where *n* is the vowel position in a word.

The fundamental frequency of vowels and words were calculated using the Speech Signal Processing Toolkit (version 3.7) [6]. The relative fundamental frequency of a vowel (*rF*) was calculated by subtracting an averaged logarithm of the fundamental frequency of the word (log *mfv*) from an averaged logarithm of the fundamental frequency of the vowel (log *mfw*) (Eq. 3).

$$(3) \qquad rF_n = \log mfv_n - \log mfw$$

where *n* is the vowel position in the word.

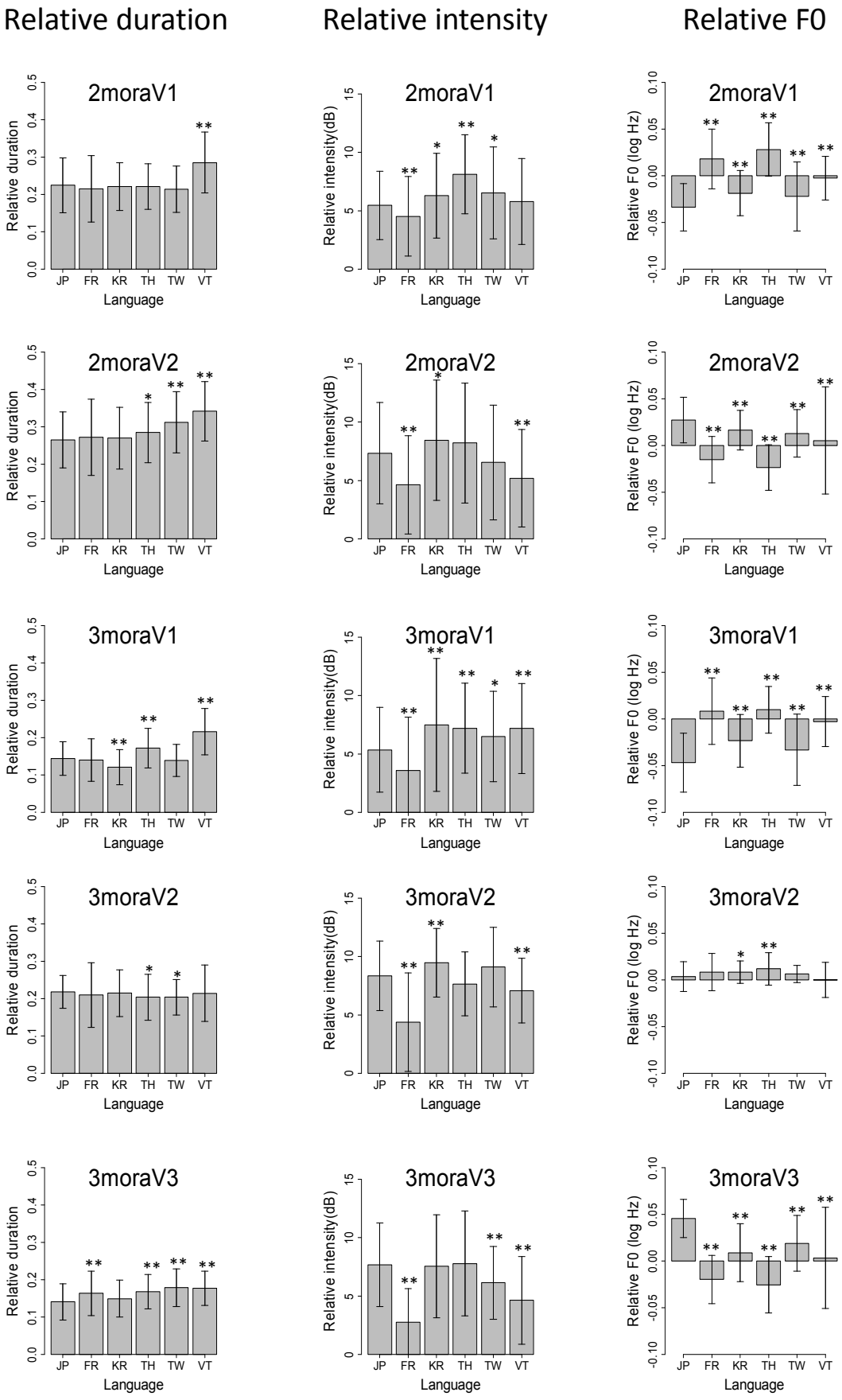
3.2. Results

The average and standard deviation of the relative duration, the relative intensity, and the relative fundamental frequency are respectively shown in the left, center, and right column of Figure 1 for each vowel position in each word length for each language. Welch's t test was conducted to examine the difference between the Japanese natives and non-natives. Significant difference from Japanese natives is indicated with asterisks in Figure 1.

4. DISCUSSION

The current results reveal that speakers in each language have different tendencies in the relative duration, intensity, and fundamental frequency in speaking Japanese words.

Figure 1: Relative duration, intensity, and fundamental frequency of the initial to final vowel in 2 or 3 mora long words produced by Japanese (JP), French (FR), Thai (TH), Taiwanese (TW), and Vietnamese (VT) speakers. Significant difference from Japanese speakers is indicated by asterisks (*: $p < .05$, **: $p < .01$).



For example, Thai and Vietnamese speakers had tendencies for the relative duration to be longer for first and last vowels than Japanese speakers. Taiwanese speakers had longer relative duration for last vowels. The results suggest that Thai, Vietnamese, and Taiwanese speakers do not properly control vowel duration in Japanese speech.

French speakers had lower relative intensity for all vowels, suggesting that they pronounce vowels more weakly but consonants more strongly than Japanese speakers. In other words, French speakers speak Japanese with less intensity difference between vowels and consonants than Japanese speakers. Korean speakers had a higher relative intensity tendency for all vowels, This suggests that they pronounce vowels more strongly but consonants more weakly than Japanese speakers. In other words, Korean speakers speak Japanese with greater intensity difference between vowels and consonants than Japanese speakers. Thai and Taiwanese speakers had a higher relative intensity for the first vowel, and Vietnamese speakers had a lower relative intensity for the last vowel. These results suggest that all non-native speakers do not properly control vowel intensity in Japanese speech.

Compared to Japanese speakers, all non-native speakers had higher relative fundamental frequency for first vowels, but lower relative fundamental frequency for last vowels. In particular, French and Thai speakers had an opposite sign of relative fundamental frequency, suggesting that they have an inverted pattern in fundamental frequency. The current results indicate that all non-native speakers do not properly control fundamental frequency in Japanese speech. Although Thai, Taiwanese, and Vietnamese are tone languages, the current results reveal that speakers of these languages are not good at using pitch accent in Japanese.

These differences in acoustic features probably depend on the characteristics of speakers' native languages. However, it is unclear which characteristics of their languages caused the difference from Japanese speakers. To clarify how their native language affects their Japanese pronunciation, further research is necessary into their native language features. In addition, it is desirable to increase variation in accent pattern, speaking rate, native language, and analysis object such as consonants. With these analyses of acoustic features, the perception of naturalness or "goodness" of their Japanese pronunciation should be studied in future.

4. ACKNOWLEDGEMENTS

This study was supported by JSPS KAKENHI Grant Numbers 25284080 and 26370464 and by a cooperative research grant (2013-2014) and a specified research grant (2013-2014, 2015-2016) of Aichi Shukutoku University. We would like to thank Professor Kyung-Ja Ryoo and Lecturer Kazuya Iihoshi of Sangmyung University, Professor Luong Chi Mai of Vietnam Academy of Science and Technology, Professors Nguyen Thu Huong and Nguyen Tien Luc of Vietnam National University - Ho Chi Minh City, Professor Chiu-yu Tseng of Institute of Linguistics, Academia Sinica, Professors Chang-Ho Lin and Meng-Ling Hsu of Ming Chuan University, Professor Rong-Kuan Shen of Shih Hsin University, Professor Yaw-Huei Maa of Tamkang University, Dr. Chatchawarn Hansakunbuntheung of National Electronics and Computer Technology Center, Professor Kanokwan Atchariyachanvanich of King Mongkut's Institute of Technology Ladkrabang, Professor Nagul Cooharojananone of Chulalongkorn University, and Dr. Takaaki Shochi of University of Bordeaux for their assistance in the utterance recordings.

5. REFERENCES

[1] Amano, S., Yamakawa, K., and Kondo, M. 2012. Discriminant variables for Japanese speeches produced by Japanese and Chinese natives. Proc. Spring Meeting Acoustical Society of Japan, 409-410. (In Japanese).

[2] Amano, S., and Yamakawa, K. 2013. The use of durational variables to characterize the rhythmic patterns of non-fluent Japanese utterance by non-native speakers. J. Acoust. Soc. Am. 134, 4246.

[3] Amino, K., and Osanai, T. 2011. Realisation of the prosodic structure of spoken telephone numbers by native and non-native speakers of Japanese. Proceedings of ICPhS, 2011, 236-239.

[4] Kato, H., Tsuzaki, M., and Sagisaka, Y. 2002. Effects of phoneme class and duration on the acceptability of temporal modification in speech. J. Acoust. Soc. Am. 111, 387-400.

[5] Kondo, M. 2012. L1-specific and language-universal interference on phonological acquisition of Japanese learners: Results of questionnaires to Japanese teachers. Bulletin of the Graduate Division of Letters, Arts and Sciences of Waseda University, III, 57, 21-34 (in Japanese).

[6] Speech Signal Processing Toolkit (SPTK ver. 3.7) http://sp-tk.sourceforge.net/

[7] Sato, T. 1995. A comparison of phonemes and prosody in the evaluation of spoken Japanese. Japanese Language Education around the Globe, 5, 139-154. (in Japanese).

[8] Sukegawa, Y. 1993. Utterance tendency of non-native Japanese speakers. Japanese Speech and Education, Research Report of Grant-in-Aid for Scientific

Research on Priority Areas by Ministry of Education, Science and Culture, 187-244 (in Japanese).

[9] Yamakawa, K., and Amano, S. 2014. Perceptual discrimination of Japanese utterances of native and non-native Japanese speakers. Bulletin of Aichi Shukutoku University, Faculty of Human Informatics (4), 15-19. (in Japanese).