IDENTIFYING ACOUSTIC FEATURES THAT CAUSE UNNATURALNESS OF NON-NATIVE SPEAKERS' JAPANESE

Kimiko Yamakawa[1], Shigeaki Amano[2]

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

ABSTRACT

Non-native Japanese speakers' pronunciation of Japanese often sounds unnatural to native Japanese speakers. The aim of this study is to specify the acoustic features that cause unnaturalness in non-native Japanese speakers' pronunciation. Japanese sentences spoken by Taiwan Mandarin, French, Korean, Thai, Vietnamese, and Japanese speakers were extracted from the non-native speakers' read-speech database. Multiple regression analyses were performed with each sentence's naturalness rating score as a dependent variable, and the acoustic features of vowels in the sentence as independent variables. On the basis of standardized partial regression coefficients, it was found that unnaturalness is related to vowel duration in Taiwan Mandarin and Vietnamese speakers, and to a pattern of fundamental frequency as well as vowel location in French, Korean, and Thai speakers. It is suggested that unnaturalness depends on prosodic characteristics of a non-native speaker's first language, such as fundamental frequency and duration.

Keywords: unnaturalness, non-native speaker of Japanese, acoustic features.

1. INTRODUCTION

As of 2015, it was estimated that there were about 3.65 million non-native speakers learning the Japanese language [1], which means there is a great demand for efficient and practical Japanese education. Speaking Japanese is one of the educational fields in which such demands are increasing. In the early stages of Japanese learning, non-native speakers make many and various speech errors [3, 5]. The number of speech errors decreases, and the amount of fluency increases as they progress in learning Japanese. However, unnaturalness remains in non-native speakers' pronunciations, even if they have developed advanced Japanese skills and correctly pronounce a phoneme sequence of Japanese words [7]. Non-native Japanese speakers' unnatural pronunciation may hinder smooth communication with native Japanese speakers [6]. Therefore, it is important for non-native speakers to acquire the skills to speak Japanese naturally.

However, an efficient and practical method for teaching natural Japanese pronunciation has not yet been developed. This is likely because the acoustic features that are related to the unnaturalness of pronunciation have not been determined. In addition, the acoustic features of unnaturalness might be uncommon and vary between the first languages of non-native speakers. Acoustic features of Japanese pronunciations by non-native speakers have differing tendencies in duration, intensity, and fundamental frequency based on the speaker's the first language [8]. Thus, the notion of language dependency is probable.

With that background in mind, this study aimed to specify acoustic features that cause unnaturalness of Japanese pronunciations by non-native Japanese speakers of various first languages.

2. EVALUATION EXPERIMENT

2.1 Stimuli

For the stimuli of the evaluation experiments, Japanese sentences spoken by 10 native Japanese speakers and 60 non-native Japanese speakers (10 each of French, Korean, Taiwanese Mandarin, Thai, and Vietnamese) were extracted from the non-native speakers' read-speech database [9]. These first languages of the non-native speakers were selected because they have phonological characteristics different from Japanese. For example, the rhythmic unit of all these languages is a syllable whereas the unit of Japanese language is a mora [2, 10]. French and Korean languages have a phrase accent and Thai, Taiwan Mandarin, and Vietnamese languages have a tone accent whereas Japanese language has a pitch accent. The Japanese proficiency of the non-native Japanese speakers was not controlled, but they could at least read Hiragana (Japanese moraic orthography) and had the ability to understand basic Japanese that corresponded to N4 level in the Japanese-language proficiency test (JLPT).

The stimuli comprised 29 Japanese words embedded in the carrier sentence, /korewa __ dato omoi masu/ ("I suppose that this is __"). These Japanese words were two or three mora long with no accent, and did not contain special moras such as long vowels, geminate consonants, or moraic nasals. A total of 1,740 stimuli (29 words x 6 languages x 10 speakers) were used for the experiment, and 60 stimuli (1 word x 6 languages x 10 speakers) were used for the practice session.

2.2 Participants

Twenty-two monolingual native speakers of Japanese (2 males and 20 females) with normal hearing ability participated in the evaluation experiment. Their average age was 27.9 years (Min=20, Max=35, SD=6.2). They were paid for their participation.

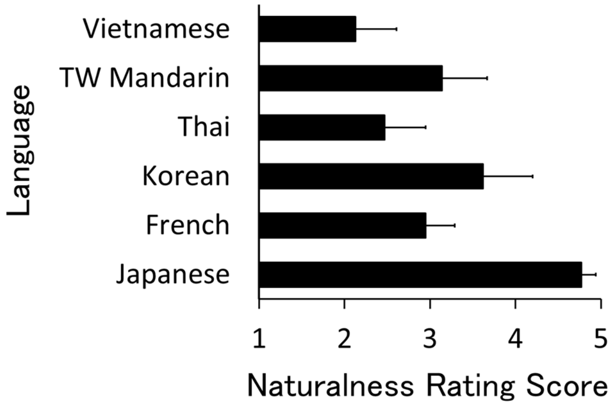
2.3 Procedure

The experiment was conducted in a soundproof room at Aichi Shukutoku University. The stimuli were presented to the participants through headphones (Sony MDR-Z900HD) via an audio interface (Roland DUO-CAPTURE EX) connected to a laptop computer (Toshiba SS-RX2). Immediately after the stimulus presentation, the target word was displayed on a computer screen in Japanese Hiragana orthography. Along with the target word, there were also shown two response buttons (correct / incorrect) for correctness judgement, and a 5-point rating scale for naturalness evaluation (with 1 equaling unnatural and 5 equaling natural).

By clicking one of the two response buttons, the participants judged whether a presented stimulus was pronounced with a correct phoneme sequence as the target word displayed on the screen. This judgement was aimed to select correctly pronounced items for the following analysis in Section 3.

Then, the participants evaluated an impression of the presented stimulus in terms of naturally spoken Japanese by clicking one of the five numbers on the 5-point rating scale (1: very unnatural - 5: very natural). The 1,740 stimuli were presented in a randomized order for each participant.

Figure 1: Naturalness rating score as a function of speaker's first language.



2.4 Results

Stimuli were selected that all participants had judged as being correctly pronounced. For each of the stimuli, the average naturalness rating score was calculated. Figure 1 shows the mean of the naturalness score for each first language of speakers.

3. ANALYSIS

3.1. Acoustic features

Four acoustic features described below were used as variables in the analysis.

The duration of a vowel and sentence was calculated by subtracting their start time from their end time. The relative duration of a vowel (*rD*) was calculated by dividing the vowel's duration (*Dv*) by the sentence duration (*Ds*) (Eq. 1).

$$rD_n = Dv_n \,/\, Ds \qquad (1)$$

where *n* is the vowel's position in the sentence.

Using a rectangular window of 6-ms width with a 1-ms shift, the intensity of a vowel and sentence was calculated 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 sentence's averaged intensity (*mIs*) from the vowel's averaged intensity (*mIv*) (Eq.2).

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

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

The fundamental frequency of vowels and sentences was calculated using the STRAIGHT system [4]. The relative fundamental frequency of vowels (*rF*) was calculated by subtracting an averaged logarithm of the fundamental frequency of the sentence (*m*log *fs*) from an averaged logarithm of the fundamental frequency of the vowel (*m*log *fv*) (Eq. 3).

$$rF_n = \; m \log f v_n - m \log f s \qquad (3)$$

where n is the vowel position in the sentence.

The Voice Center Time (VCT) is defined as the midpoint time of the start and end times of a vowel. The relative VCT (*rVCT*) was calculated by dividing *VCT* by *Ds* (Eq. 4).

$$rVCT_n = VCT_n \,/\, Ds \qquad (4)$$

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

3.2 Procedure

The non-native speakers' stimuli that were judged as correct pronunciations in the evaluation experiment were used as analysis objects. The native speakers' items that had a higher naturalness score than the average of the native speakers' scores were also used for the analysis. Using this data of native Japanese speakers, an average of each acoustic feature described in Section 3.1 was calculated for each vowel in each item. This average was regarded as the reference point of naturally spoken Japanese. Deviation of each item from the reference point was obtained as a logarithmic converted sum of the root mean square error.

Using the deviation values of acoustic features as independent variables, and the naturalness rating score as a dependent variable, multiple regression analysis was performed with a paired data set of native speakers and non-native speakers in one of the first languages.

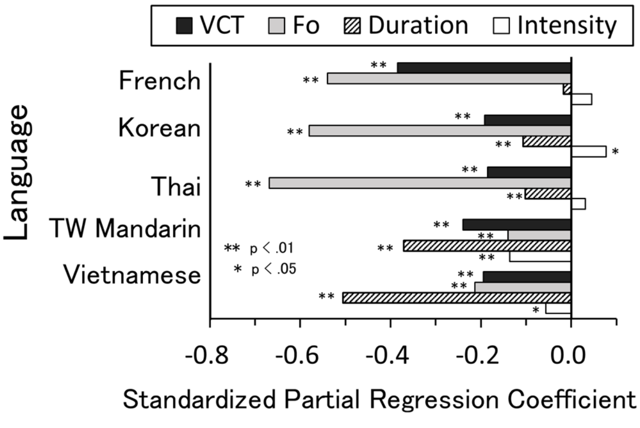
3.3 Results

Figure 2 shows the standardized partial regression coefficient obtained by the multiple regression analysis. The adjusted coefficient of determination (adjusted R^2) was 0.609, 0.443, 0.650, 0.466, and 0.772 for French, Korean, Thai, Taiwan Mandarin, and Vietnamese speakers, respectively. When all data of non-native speakers was used, adjusted R^2 was 0.628. These results indicate that regression is fairly successful with the variables used in this study, and that unnaturalness of speech is related to these variables.

4. DISCUSSION

The results of this study show that unnaturalness of Japanese pronunciation by non-native speakers is related to acoustic features such as fundamental frequency and durations, but the effective acoustic features are different between the first languages of non-native speakers.

Figure 2: Standardized partial regression coefficient as a function of speaker's first language.



For French, Korean, and Thai speakers, absolute value of standardized partial regression coefficients of fundamental frequency was bigger than other variables (Figure 2). Hence, it can be said that their unnaturalness is caused by fundamental frequency deviation from the natural fundamental frequency pattern of Japanese speakers. In fact, French, Korean and Thai speakers tended to pronounce word items with an initial-high accent pattern (i.e., high-low-low), whereas Japanese speakers pronounced the word items with a flat accent pattern (i.e., low-high-high). This fundamental frequency difference in accent patterns would be one cause of unnaturalness.

Standardized partial regression coefficients indicate that VCT contributes to French speakers' unnaturalness. Because the VCT represents the vowel's time position unnaturalness among French speakers is probably caused by a deviation from the natural rhythmic pattern of Japanese speakers.

For Taiwan Mandarin and Vietnamese speakers, standardized partial regression coefficients indicate that vowel duration contributes to the unnaturalness of their pronunciation. Taiwan Mandarin speakers tended to extend the second from the last vowel in a sentence longer than Japanese speakers. Vietnamese speakers tended to pronounce every vowel longer than Japanese speakers. These durational deviations would be one cause of the unnaturalness of Taiwan Mandarin and Vietnamese speakers' pronunciation.

The French and Korean languages have a phrase accent that differs from a mora accent in Japanese. Although both accents are realized by fundamental frequency control, the accent unit is different. This difference suggests that French and Korean speakers are not good at controlling their fundamental frequency pattern in Japanese. For this reason, French and Korean speakers' unnaturalness depends on fundamental frequency, as observed in this study.

French does not have a distinction between short and long vowels, which suggests that French speakers are not good at durational control. Because of this, French speakers' unnaturalness depends on duration as well as fundamental frequency, as observed in this study.

Taiwan Mandarin and Vietnamese have a tone accent, but do not have distinctions between short and long vowels. The tone accent is realized by fundamental frequency control, and vowel length distinction is realized by duration control. Hence, Taiwan Mandarin and Vietnamese speakers are probably good at controlling fundamental frequency pattern, but not duration. Therefore, Taiwan

Mandarin and Vietnamese speakers' unnaturalness depends on duration, as observed in this study.

The Thai language has a tone accent and distinguishes between short and long vowels. This suggests that Thai speakers are probably good at controlling both fundamental frequency and duration, and that their unnaturalness is not related to these two features. However, the results of this study showed that Thai speakers' unnaturalness depends on fundamental frequency. The cause of this contradiction is uncertain, and unknown factors may affect their unnaturalness. This should be studied in future research.

Although some unknown factors may exist, the results of this study suggest that the unnaturalness of non-native speakers depends on prosodic characteristics of their first language in terms of fundamental frequency and duration.

The adjusted R^2 was not as high for Taiwan Mandarin and Korean speakers as it was for speakers of the other languages in this study. This suggests that variables of acoustic features that were not used in this study may affect the unnaturalness of their speech. This possibility should be examined in a future study.

This study did not address the Japanese special mora such as geminate consonants, moraic nasals, and long vowels. Because non-native Japanese speakers are not good at pronouncing the special mora [3, 5] unnaturalness may be related to the acoustic features of the special mora. The distinction between a special mora and a normal mora is indicated by a durational difference. Therefore, duration of the special mora probably contributes to unnaturalness of non-native speakers' pronunciation. This prediction should be studied in the future.

5. ACKNOWLEDGEMENTS

This study was supported by JSPS KAKENHI Grant Numbers JP26370464, JP17K02705.

6. REFERENCES

[1] Japan Foundation. Survey report on Japanese-language education abroad 2015. https://www.jpf.go.jp/e/project/japanese/survey/result/survey15.html.

[2] Han, M. 1994. Acoustic manifestations of mora timing in Japanese," The Journal of the Acoustical Society of America 96(1), 73-82.

[3] 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).

[4] STRAIGHT (version V40_005b), http://www.wakayama-.ac.jp/~kawahara/puzzlet/STRAIGHTtipse/

[5] 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).

[6] Thomas, J. 1983. Cross-cultural pragmatic failure. Applied Linguistics 4, 91-112, 1983.

[7] Yamakawa, K., 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).

[8] Yamakawa K., Amano, S. 2016. Acoustic feature representing the unnaturalness of Japanese spoken by non-native speakers. International Symposium on the Acquisition of Second Language Speech (New Sounds 2016) P-II-17.

[9] Yamakawa K., Amano, S., Kondo, M. 2014. Development of Japanese read-word database for non-native speakers of Japanese. Proceedings of 17th Oriental Chapter of the International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques (Oriental COCOSDA), 65-70.

[10] Vance, T. J. 2008. *The Sounds of Japanese*. Cambridge: Cambridge University Press.