

Impact of Acoustic Similarity on Perceptual Assimilation of English schwa by Native Speakers of Japanese*

© Kanako Tomaru (Sophia Univ.), Takayuki Arai (Sophia Univ.)

1 Background

1.1 Perceptual assimilation

When people hear non-native speech sounds, they tend to perceptually classify these unknown sounds under their native speech sound categories. For example, English liquids, i.e., /r/ and /l/, are often perceptually categorized as variant of the Japanese flap, i.e., /ɾ/, by native speakers of Japanese. This process, which is called “perceptual assimilation (PA),” was introduced in the Perceptual Assimilation Model (PAM) by Best (1991, 1995) [1, 2]. The PAM model explains the perception of non-native sounds in terms of articulatory similarity to native segments. In some cases, phonemically contrastive non-native sounds lose their contrast when assimilated¹. This is the case of the PA of English liquids to Japanese /ɾ/ (*Single-Category; SC type*). In other cases, the members of a non-native phoneme set retain their contrast when assimilated to different native categories (*Two-Category; TC type*). For example, English vowels /i/ and /u/ are assimilated to Japanese /i/ and /u/ categories even though precise articulatory gestures for the English vowels are different from those for the Japanese ones [3, 4]. Whether a non-native sound is a good example can be assessed by the degree of Category- Goodness (*CG type*). It is also possible that some non-native sounds are so dissimilar from the available native sounds that they are perceived as non-speech sounds (*Non - Assimilated; NA type*).

1.2 Impact of acoustic similarity

Since listeners assess articulatory similarities between native and non-native sounds, the PA pattern is predictable to some extent based on acoustic similarities. However, several other

factors seem to play important roles. For example, although the PA of English vowels by Japanese listeners is predictable based on their “phonetic” similarities, such as the backness and roundness of the non-native vowel to be assimilated, the PA patterns of some vowels vary depending on the speaker and condition, as well as a combination of the two variables. In addition, lexical knowledge may also play a substantial role in deciding which sound category a non-native sound in question should be categorized [5]. Tomaru (2009) suggests that PA patterns are affected by lexical well-formedness in a listener’s native language. For example, some English vowels can be assimilated to more than one Japanese vowel category [3, 4]. In such cases, Japanese listeners perceptually prefer a category that can create a real word from a presented non-native stimulus. For instance, when presented with an English pseudo-word stimulus /pɒbʊ/ whose vowel can be assimilated to either /a/ or /o/, Japanese listeners prefer hearing it as /pabʊ/, which is an existing loanword for “pub,” as opposed to hearing it as /pobʊ/, which is a pseudo word in Japanese.

2 Purpose

The purpose of the present study is to investigate the impact of acoustic similarities on the PA of English schwa vowels by Japanese listeners. Although this topic has been previously discussed by Tomaru and Arai (2012, 2015) [6, 7]², several questions are remained to be unanswered.

In Tomaru and Arai’s first report (2012) [6], it was suggested that the English schwa was the most likely to be assimilated to Japanese /a/ (96%) regardless of its acoustic similarity to Japanese /u/ on the equivalent rectangular bandwidth scale. However, the impact of acoustic similarities

*音響的な類似が日本語母語話者による英語の schwa の知覚同化へ与える影響, 渡丸嘉菜子(上智大・理工), 荒井隆行 (上智大・理工)

Table 1. Average distance in F1–F0 and F3–F2. Values less than or more than 3 Bark are displayed separately. Numbers of analyzed vowel tokens are in parentheses.

Vowels	<i>F1 – F0</i> (Av.)		<i>F3 – F2</i> (Av.)	
	< 3 Bark	> 3 Bark	< 3 Bark	> 3 Bark
English schwa	2.4 (1)	4.2 (31)	2.8 (2)	4.1 (30)
Japanese /a/	2.8 (5)	4.7 (963)	2.7 (34)	4.5 (934)
Japanese /u/	2.4 (251)	3.9 (74)	2.1 (113)	4.2 (212)

between vowels was not completely denied. In their report, Japanese vowels to be compared with schwa vowels were carefully spoken in experimental settings. On the other hand, listeners may recall natural spontaneous speech rather than careful speech during the PA. If so, the similarity between the English schwa and the Japanese /a/ and /u/ in spontaneous speech should be examined.

This point was investigated in their later study [7]. Compared with spontaneous speech in the first (F1) and second (F2) formant dimensions in Hz, schwa vowels were more likely to be categorized as Japanese /a/ (76%) than as /u/ (24%).

Previous studies have suggested that Japanese listeners' PA patterns may be predicted by acoustic similarity. However, a gap remains between perceptual response, i.e., a schwa is heard as a Japanese /a/ 96% of the time, and the prediction that can be made based on acoustic similarity, i.e., 76% of schwa vowels are categorized as /a/. Therefore, in the present study, the relationship between acoustic similarities and the PA pattern is further investigated.

3 Method

3.1 Materials

A total of 32 English schwa materials, the same as those employed in previous studies by Tomaru and Arai (2012, 2015), were analyzed in the present study. The speaker of these schwa vowels was a male native speaker of American English. The list of pseudo words spoken by the speaker is presented in Tomaru and Arai (2015). The Japanese vowel materials were taken from the

Corpus of Spontaneous Japanese (CSJ) [9, 10]. As in previous research, tokens of the same speaker, i.e., speaker A01M0015, were analyzed.

3.2 Analysis

The fundamental frequency (F0) values and the first three formant frequency values (F1, F2 and F3) were measured at the midpoints of a vowel's steady state in Hz. All Japanese /a/ (labeled as "a") and /u/ (labeled as "u") tokens in the speaker's recording were analyzed by using the Praat software [11]. Excluding those that had missing values, the total number of /a/ and /u/ tokens analyzed were 968 and 325, respectively.

For analysis, the frequency values of the English schwa vowels and the Japanese /a/ and /u/ vowels in Hz were converted to the Bark scale, and then transformed to measures that normalize for vocal tract (speaker) differences: F1–F0 and F3–F2 in Bark [12, 13]. The former corresponds to vowel "height," while the latter corresponds to vowel "place." High vowels show a less than 3 Bark difference between F0 and F1. On the other hand, for front vowels, F3–F2 show less than 3 Bark [12].

3.3 General comparison

The average values of F1–F0 and F3–F2 is shown in Table 1. The height and backness of vowels can be assessed based on this distance: if the distance in these dimensions is less than 3 Bark, the vowel is "high" and "front." As shown in Table 1, the English schwa vowels and the two Japanese vowels seem to be similar in terms of average distance between F1–F0 and F3–F2. However, the distribution of each group of vowels is different.

Table 2. Average Mahalanobis distance between the English schwa vowels and the Japanese /a/ and /u/ in F0-F1 and F3-F2 dimensions. The percentage of schwa categorization is shown in parentheses.

Average Mahalanobis distance to schwa vowels	
Japanese /a/	1.7 (84.4%)
Japanese /u/	3.9 (15.6%)

This difference is more apparent in a scatter plot of the English schwa vowels and the two Japanese vowels in the F1–F0 and F3–F2 dimensions (Fig. 1). Japanese /a/ and /u/ are well-differentiated in the F1–F0 dimension. Although the Japanese /u/ is more central than the English /u/, it is reasonable that the Japanese /a/ and /u/ are separated from each other in terms of backness. Schwa vowels are distributed over the Japanese /a/ and the /u/ dimensions.

3.4 Discrimination analysis

To further investigate whether the PA pattern of English schwa vowels can be predicted by acoustic similarities, we conducted a discrimination analysis using the Mahalanobis distance in the F0–F1 and F3–F2 dimensions. The results of the analysis are summarized in Table 2, which shows the average Mahalanobis distance between the schwa and the two Japanese vowels. The percentage of discrimination based on vowel distance in the F0–F1 and F3–F2 dimensions is shown in parentheses. The results indicate that when transformed to measures of distance, 84.4% of the English schwa tokens were categorized as Japanese /a/, while only 15.6% of the tokens were categorized as /u/.

4 Summary and Discussion

The present study shows that more schwa tokens were categorized as Japanese /a/ (84.4%) than in the previous report by Tomaru and Arai (2015) (76%) [7], suggesting that speaker normalized acoustic similarities between the English schwa vowel and Japanese vowels are key for predicting the PA patterns of Japanese listeners'. However, it should be noted that perceptual data showed that 96% of the same

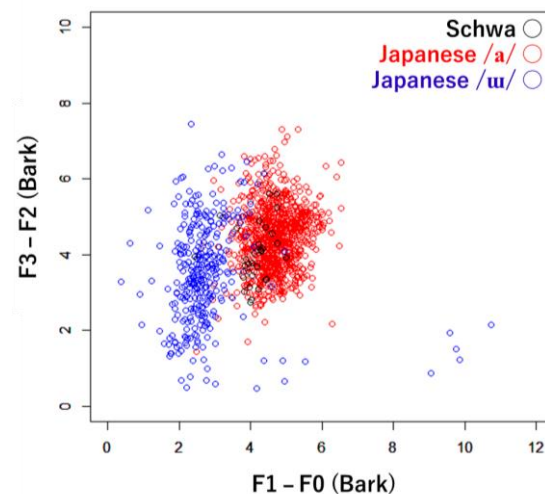


Fig 1. Scatter plot of English schwa vowels and Japanese /a/ and /u/ in the F1–F0 and F3–F2 dimensions.

schwa tokens analyzed in both the present and former studies were categorized as Japanese /a/ by native speakers of Japanese. Therefore, even though the results of the present study suggest strong impact of acoustic information on PA, other factors may also play important roles.

One of these possible factors could be the relative appearance probability of native vowels. Although the present analysis is limited to one speaker and only vowels taken from the CSJ, the vowel /a/ appeared more frequently than the vowel /u/: 1125 times vs. 530 times in one recording. Therefore, Japanese listeners may adopt more frequently used vowel categories to ambiguous native sounds. Further research should be conducted to bridge the gap between perceptual data and acoustically predicted categorization patterns.

Acknowledgements

This research was supported by the Japan Society for Promotion of Science (Kakenhi Grant Number: 16K16832).

References

- [1] Best, Haskins Laboratories Status Report on Speech Research, SR-107/108, 1–30, 1991.
- [2] Best, Speech perception and linguistic experience: Issues in cross-language research (pp. 171–204). Timonium, 1995.
- [3] Strange *et al.*, Journal of Phonetics, 26, 311–344, 1998.
- [4] Strange *et al.*, J. Acoust. Soc. Am., 109 (4), 1691–1704, 2001.
- [5] Tomaru, The 24th Meeting of the Sophia University Linguistic Society, 24, 1–20, 2009.
- [6] Tomaru & Arai, Proc. of the 26th General Meeting of the Phon. Soc. Jpn., 79–84, 2012.

¹ In this paper, the words “assimilate” and “assimilation” are used exclusively to refer to the perception-based categorization of non-native sounds into native-sound categories as in PAM. No phenomenon related to *assimilation* in

- [7] Tomaru & Arai, Proc. Autumn Meet. Acoust. Soc. Jpn., 323–326, 2015.
- [8] Tomaru, undergraduate thesis submitted to Doshisha Univ., 2008.
- [9] Maekawa, Proc. IEEE Workshop on Spontaneous Speech Processing and Recognition, 7–12, 2003.
- [10] Maekawa et al., Proc International Symposium on Largescale Knowledge Resources, 19–24, 2004.
- [11] Boersma & Weenink, Praat: doing phonetics by computer [Computer program]. Version 6.0.23, retrieved 12 December 2016 from <http://www.praat.org/>
- [12] Syrdal & Gopal, J. Acoust. Soc. Am., 79 (4), 1086–1100, 1986.
- [13] Bohn and Munro, *Language Learning & Language Teaching vol. 17 Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, John Benjamins Publishing, 2007.

articulatory phonetics are implied.

² The experimental results and discussion related to this topic were originally reported in an undergraduate thesis submitted to Doshisha Univ. in 2008 by the first author [8].