# Acoustic comparison between English schwa and Japanese vowels in spontaneous speech in terms of low-order formant frequencies *

◎Kanako Tomaru, Takayuki Arai (Sophia Univ.)

## 1 Introduction

Speech segments in a given language are usually difficult for non-native speakers to distinguish. One of the reasons is that these foreign sounds have different qualities from native sounds. This quality difference can often be an obstacle when adopting words from one language community to another. When foreign words were adopted in to a language, they may be pronounced according to the orthography of the source language; other times, the native pronunciation of a source word is, most of the time, amended to fit the sound system of the borrowing language. For example, the English word *something* /sʌmθɪŋ/ is pronounced as サムシング /samuɕingu/ when adapted to Japanese. The biggest amendment here other than the insertion of epenthetic vowels is that /θ/ turns into /ɕ/. This is necessary because there is no consonant in the Japanese consonant system that has the same quality as English /θ/.

An analogous phenomenon is seen in the perception of non-native segments. For example, Best's Perceptual Assimilation Model (PAM) predicts that listeners will categorize a heard foreign vowel sound as some vowel in their native language, for the same reason that they express a foreign segment by means of a native segment when adapting a loanword: for instance, English vowel /ʌ/ may be perceived as native Japanese vowel /a/ by a native speaker of Japanese.

This kind of adaptation of a non-native segment into a native segment is called *perceptual assimilation* [1]. Best's PAM [1] explains the perception of non-native segmental contrasts by referring to vowel categories and the articulatory similarity of foreign segments to native segments. Her claim is that the non-native phones are perceived in terms of articulatory similarity to native segments. A foreign segment that is similar to a particular native segment is expected to be placed in the same category as that native sound; this case is called *Categorized*. This process is called perceptual assimilation [1]. Subsequently, the assimilated non-native segment will be judged as either a good or a poor exemplar of the native segment; this judgment process is called *goodness rate*. A phone dissimilar to any available native sound, on the other hand, will not be assimilated to any of the native categories (leaving it *Uncategorized*). The model also predicts cases in which the segment will not be recognized as a speech sound at all, although they will be rare (*Non-assimilated*).

## 2 Purpose of the study

Tomaru and Arai [2] investigated perceptual assimilation of English schwa by native speakers of Japanese, for the first time. They found that English schwa was perceptually assimilated to the Japanese vowel /a/ (96%). However, their formant analysis contradicted the perceptual results. That is, it showed that the English schwa used for their perceptual experiment was more similar to the Japanese /ɯ/ vowel than the Japanese /a/ in terms of the first and second formant frequencies (F1 and F2).

In that study, Japanese vowels to be compared with English schwa were recorded in experimental settings. However, vowel quality in ordinary conversational speech may diverge from that in carefully read speech [3]. Therefore, in this study, we conducted a formant analysis of Japanese /ɯ/ and /a/ in spontaneous speech to investigate which of the vowels is acoustically closer to English schwa. Since former studies report that the formant values shows

*英語の弱化母音と自発的に発話された日本語母音との音響的比較―低次フォルマント周波数の観点から―,渡丸嘉菜子, 荒井隆行(上智大・理工).

Table. 1 Basic comparison among the three vowel categories: English schwa, Japanese /a/ and Japanese /ɯ/.

| | F1 in Hz (Av.) | F2 in Hz (Av.) | Number of analyzed tokens | Speaker's sex | Speech type |
|---|---|---|---|---|---|
| English schwa | 606 | 1333 | 55 | male | read |
| Japanese /a/ | 681 | 1252 | 200 | male | spontaneous monologue |
| Japanese /ɯ/ | 439 | 1582 | 200 | male | spontaneous monologue |

tendency to reduce, or in other words, centered [4-6], we expect the Japanese /a/ to be reduced so that its F1 and F2 get closer to those of English schwa.

## 3 Materials

The English schwa materials analyzed in this research were the same materials used for the perceptual experiment in the preceding research discussed above [2]. These schwa vowels were produced by a male native speaker of American English. In recording, he read a list of nonsense words that contained schwa in different positions: word initially, medially and finally (see Appendix). The recording took place in a quiet room.

The Japanese vowel materials form a small subset of the Corpus of Spontaneous Japanese (CSJ) [7, 8]. This corpus consists of Japanese spontaneous monologues such as academic presentations [3, 7, 8].

## 4 Analysis

For the present analysis, we picked one male speaker from CSJ (speaker ID: A01M0015) who sounded to be around the age of the American speaker. We selected 264 tokens of the vowel /a/ and 402 tokens of the vowel /ɯ/ from his utterances. However, because of devoicing or lack of sustained steady state, 64 tokens of /a/ and 202 tokens of /ɯ/ had to be eliminated from the analysis. Thus, ultimately, 200 tokens of each vowel were analyzed.

### 4.1 Overall Results

The results of the basic comparison among the three vowel categories is summarized in Table 1, which shows the average of F1 and F2 values, the total number of analyzed materials, speaker's sex, and the type of speech. A quick glance at the table might tell you that English schwa vowels in read speech have formant frequencies closer to Japanese /a/ vowels in spontaneous speech than spontaneous Japanese /ɯ/ vowels in spontaneous speech.

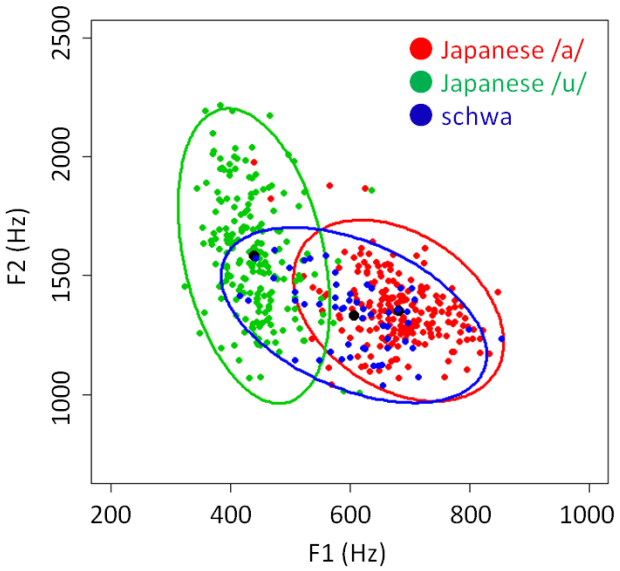


Fig. 1 Scatter plot of Japanese /a/ (red), Japanese /ɯ/ (green), and English schwa (blue).

Figure 1 shows the distribution of the three vowel categories in question. The scatter plot for Japanese /a/ is shown in red, that for /ɯ/ in green, and that for schwa in blue. A 95% confidence ellipse is also shown for each vowel category. The black dot in each vowel space is the center of the confidence ellipse. When you take a look at the vowel spaces in Fig. 1, it seems that English schwa share more space with Japanese /a/ than with Japanese /ɯ/. Thus, the results in general may suggest that English schwa is acoustically closer to Japanese /a/ than to /ɯ/.

### 4.2 Cluster analysis

In order to further investigate acoustic similarity between English schwa and the two Japanese vowels, we next conducted a discrimination analysis using the Mahalanobis distance. The results of the analysis are summarized in Table 2, which shows the average Mahalanobis distance between English schwa tokens and the Japanese /a/ and /ɯ/ distributions. The

Table. 2 Average Mahalanobis distance between English schwa vowels and Japanese /a/ and /ɯ/. The percentage of categorization of schwa in each of the categories is also shown.

| | Mahalanobis distance (Av.) | Percentage of categorization |
|---|---|---|
| Japanese /a/ | 4.08 | 76% |
| Japanese /ɯ/ | 13.98 | 24% |

second column also shows the percentage of categorizations in each of the two Japanese vowel categories. Through this analysis, we found that 76% of the schwa tokens were best categorized as Japanese /a/ while only 24% was categorized as Japanese /ɯ/.

## 5 Discussion

As mentioned above, the purpose of the present research is to investigate the acoustic similarity between English schwa and two Japanese vowels: /a/ and /ɯ/. The motivation behind the current analysis is the previous work by Tomaru and Arai (2012) [2], who showed that English schwa vowels were predominantly perceptually assimilated to Japanese /a/, in preference to other vowels, even though schwa seemed to be similar to Japanese /ɯ/ in their acoustic analysis of read speech. However, it is likely that acoustic properties of carefully read speech differ from those of spontaneous speech. Therefore, it is reasonable to speculate that a listener's perceptual assimilation pattern might reflect the acoustic properties of the latter rather than the former. To further investigate the question, we compared English schwa tokens with spontaneously uttered Japanese vowels. Through a discrimination analysis, we found that English schwa vowels were closer to Japanese /a/ than to /ɯ/.

The results of the current study have implications for the theory of perceptual assimilation and vowel reduction in Japanese spontaneous speech.

Firstly, it is implied from our results that categorization during the perceptual assimilation is based on statistical distribution of spontaneous speech. More interestingly, the distance between the centers of distributional ellipses seems to be more important than the absolute similarity of individual tokens. As seen in Table 1, a discrimination analysis showed that 24% of English schwa tokens were categorized as Japanese /ɯ/ although the results of a perceptual experiment showed that nearly 100% of schwa vowels were assimilated to Japanese /a/. If perceptual assimilation is a process of calculating the distance between an individual token (here, a schwa) and the distribution of Japanese vowels, at least 20% of schwa vowels should have been assimilated to Japanese /ɯ/. However, this was not the case. There are two possible account for the gap between perceptual data and production data. Firstly, a listener may compare the degree of lip rounding of the vowels which is expected to be reflected in F3. Although Japanese /ɯ/ is said to be unrounded, schwa and Japanese /a/ may share the same quality in terms of roundness. Analysis in terms of F3 should be included in the future research. Second possibility is that, perceptual assimilation by Japanese listeners is based on the direction and the center of the distribution. As observed in Fig. 1, the center of the schwa distribution is closer to the center of the Japanese /a/ distribution than to that of the /ɯ/. In addition, schwa and Japanese /a/ have variation in F1 where as Japanese /ɯ/ has variation in F. This implies that, Japanese listeners have created a distributional space for English schwa and have compared its distributional center with those of Japanese vowels. Although this explanation seems largely reasonable, one question remains: how do Japanese listeners create a distributional space for schwa from only a limited number of inputs? Because the Japanese listeners who participated in the former research had relatively good command of English, it is possible that they already had a distributional space for English schwa. This issue should be taken up in the future research.

Second, the analysis of spontaneous speech in this research gives supporting evidence that Japanese vowels are somewhat reduced and centered in spontaneous speech. Table 3 summarizes the averages of the first two formant frequencies of

Table. 3 Formant frequencies of Japanese /a/ and /ɯ/ in read and spontaneous speech.

| | F1 in Hz (Av.) | F2 in Hz (Av.) | Number of analyzed tokens | Speaker's sex | Speech type |
|---|---|---|---|---|---|
| Japanese /a/ in Kasuya *et al.* (1968) [9] | 775 | 1163 | 85 | male | read |
| Japanese /a/ in the present study | 681 | 1252 | 200 | male | spontaneous monologue |
| Japanese /ɯ/ in Kasuya *et al.* (1968) [9] | 363 | 1300 | 90 | male | read |
| Japanese /ɯ/ in the present study | 439 | 1582 | 200 | male | spontaneous monologue |

Japanese /a/ and /ɯ/ spoken under different conditions: read speech (adopted from Kasuya *et al.* (1968) [9]) and spontaneous monologue (from the present study). In this table, you can see that for /a/ in spontaneous speech, F1 is reduced and F2 is increased as compared to read speech. These changes are an indication of higher tongue position and less opened mouth. On the other hand, for /ɯ/ in spontaneous speech, increased F1 may indicate that the tongue is lower than it is for /ɯ/ in read speech. Although this observation is only preliminary, these results support the previous findings on vowel variation and centralization in spontaneous speech [4-6].

## References


[1] Best, *Speech perception and linguistic experience: Issues in cross-language research* (pp.171-204). Timonium, 1995.

[2] Tomaru & Arai, Proc. of the 26th General Meeting of the Phon. Soc. of Japan, 79-84, 2012.

[3] Furui *et al.*, Speech Com., 47, 208-219, 2005.

[4] Keating & Hoffman, Phonetica, 41 (4), 191-207, 1984.

[5] Harmegnies & Poch-Olive, Speech Com., 11, 429-437, 1992.

[6] Nicolaidis, Proc. of 15th ICPhS, 1-4, 2003.

[7] Maekawa, Proc. IEEE Workshop on Spontaneous Speech Processing and Recognition, 7-12, 2003.

[8] Maelawa *et al.*, Proc. International Symposium on Largescale Knowledge Resources, 19-24, 2004.

[9] Kasuya *et al.*, J. Acoust. Soc. Japan, 24 (6), 355-364, 1968.


## Acknowledgements


We are very thankful to Professor C. Best for her comments and suggestions on our manuscript.


## Appendix

Nonsense words that contain schwa in (1) word initial, (2) word medial, and (3) word final position.

(1) Initial position

*abive*, [əˈbaiv]; *adize*, [əˈdaiz]; *aguy*, [əˈgai];*azide*, [əˈzaid]; *abeep*, [əˈbip]; *adeat*, [əˈdit]; *ageek*, [əˈgik]; *azea*, [əˈzi]; *aboof*, [əˈbuf]; *adoose*, [əˈdus]; *agooke*, [əˈguk]; *azuit*, [əˈzut].

(2) Medial position

*tababite* [ˈtæbəˌbait]; *tagagite* [ˈtægəˌgait]; *tazazite* [ˈtæzəˌzait]; *tababet* [ˈtæbəˌbɛt]; *tedadet* [ˈtædəˌdɛt]; *tegaget* [ˈtægəˌgɛt]; *tezazet* [ˈtɛzəˌzɛt]; *teababeat* [ˈtibəˌbit]; *teadadeat* [ˈtidəˌdit]; *teagageat* [ˈtigəˌgit]; *teazazeat* [ˈtizəˌzit]; *cobaboke* [ˈkoʊbəˌboʊk]; *cobaboke* [ˈkoʊbəˌboʊk]; *codadoke* [ˈkoʊdəˌdoʊk]; *cogagoke* [ˈkoʊgəˌgoʊk]; *cozazoke* [ˈkoʊzəˌzoʊk]; *cubabuke* [ˈkubəˌbuk]; *cugaguke* [ˈkugəˌguk]; *cuzazuke* [ˈkuzəˌzuk].

(3) Final position

*sabba*, [ˈsæbə]; *sadda*, [ˈsædə]; *saga*, [ˈsægə]; *sazza*, [ˈsæzə]; *keyba*, [ˈkibə]; *keyda*, [ˈkidə]; *keyga*, [ˈkigə]; *keyza*, [ˈkizə]; *pooba*, [ˈpubə]; *pooda*, [ˈpudə]; *pooga*, [ˈpugə]; *pooza*, [ˈpuzə]; *cabub*, [ˈkæbəb]; *cadud*, [ˈkædəd]; *cagug*, [ˈkægəg]; *cazuz*, [ˈkæzəz]; *pebub*, [ˈpibəb]; *pedud*, [ˈpidəd]; *pegug*, [ˈpigəg]; *pezuz*, [ˈpizəz]; *boobub*, [ˈbubəb]; *boodud*, [ˈbudəd]; *boogug*, [ˈbugəg]; *boozuz*, [ˈbuzəz].