Evaluating the articulatory similarity using formant and fundamental frequencies during perceptual assimilation of English schwa by native speakers of Japanese

Kanako Tomaru* and Takayuki Arai†

Faculty of Science and Technology, Sophia University,
7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan

(Received 18 June 2018, Accepted for publication 4 February 2019)

Abstract: The main hypothesis behind the perceptual assimilation model (PAM) is that listeners perceive non-native speech sounds based on articulatory similarities between the non-native speech sounds and native speech sounds. When non-native and native sounds are substantially similar, non-native sounds are perceived as equivalent to native ones. Former research shows that the schwa vowel is perceptually assimilated to Japanese /a/ almost exclusively. In this study, we investigate acoustical cues that Japanese listeners rely on to assess the articulatory similarities between the English schwa and the Japanese vowels during the perceptual assimilation. Traditionally, the first two formants (F1 and F2) are considered to be effective for assimilation judgements; however, former investigations imply that these dimensions may be insufficient. In this study, we compared the schwa and the five Japanese vowels in the dimensions of vowel openness and backness using additional information, i.e., the third formant (F3) and fundamental frequencies (F0). The results of the analyses suggest that Japanese listeners use F1, F2, F3 and F0 information to assess articulatory similarities between the schwa and the Japanese vowels during perceptual assimilation.

Keywords: Perceptual assimilation model (PAM), English schwa, Japanese vowels, Formant frequency, Fundamental frequency, Articulatory similarity

PACS number: 43.71.−k, 43.71.Es, 43.71.Hw, 43.71.Sy, 43.72.Ja [doi:10.1250/ast.40.233]

1. INTRODUCTION

1.1. Perceptual Assimilation Model

It is widely accepted that infants’ sensitivity to phonetic contrasts that are unnecessary to native language (L1) decreases at least before their first birthday [1]. The process of adults’ non-native speech perception and the effects of listeners’ L1 on this process have been studied and modeled by many scholars [2–4, among others]. Among such models, Best’s perceptual assimilation model (PAM) [5] focuses on the acquisition of contrast of second language (L2) sounds in the light of the articulatory similarity of speech perception as opposed to the psycho-acoustic view.

Best and Tyler [5,6] hypothesize that the difficulty of phonetic contrast perception depends on the similarity between the L1 and non-native contrasts in question. The main idea of the PAM is that non-native speech sounds are perceptually categorized, or in terms of the L2 learner, perceptually assimilated, into native sound categories if the non-native sounds are substantially similar to the native ones. The PAM suggests four main patterns of perceptual assimilation: two-category (TC), single-category (SC), category goodness (CG), and non-assimilable (NA). In TC perception, two contrastive non-native sounds are perceived as two distinct native categories. The SC pattern is the case of native speakers of Japanese perceiving English /r/–/l/ contrast: the contrast is lost and the English /r/ and /l/ are heard as the Japanese /R/ category, even though these English liquids are somewhat discrepant from the Japanese category [7]. When one of the contrastive non-native sound categories is a better example of a native category than the other, the assimilation results may show differences in CG ratings. Finally, when non-native sounds are too discrepant from native sound properties, these sounds are NA.

*e-mail: k.tomaru@mejiro.ac.jp
†Current affiliation: Faculty of Foreign Language Studies, Mejiro University, 4–31–1 Nakaokchiae, Shinjuku-ku, Tokyo, 161–8539 Japan
‡e-mail: arai@sophia.ac.jp
1.2. Focus of the Present Study

The PAM focuses on the perception of contrast in line with the development of L1 sound categories and the effect of such L1 categories on L2 perception [5,6]. Technically, PAM prediction is applicable to contrastive consonants and vowels [8]. However, the PAM hypothesis has been tested almost exclusively on consonants because (1) vowels are less likely than consonants to show categorical perception in an obvious way [9], and (2) vowels are less likely to be heard as NA to speech sounds [8]. On the other hand, it should be noted that some L2 sounds that are non-contrastive in nature must be learnt as if they are contrastive during L2 learning in order to perceptually discriminate the quality differences between L2 and L1 sounds. Thus, although the main scope of the PAM is contrastive speech sounds, perceptual patterns of L2 non-contrastive sounds should be investigated using the PAM framework.

Strictly speaking, two phones are contrastive when they make a minimal pair of words, e.g., the English words “pay” [pei] vs “bay” [beɪ]. In this sense, an English schwa vowel is not fully contrastive with other vowels because the contrastive status of English schwa vowels depends on their position in a syllable and on linguistic motivation. Generally, schwa vowels are assumed to be “reduced vowels” that occur in the unstressed position of a syllable [10,11]. Some schwa vowels are unstressed alternatives of stressed full vowels, e.g., “atom” [ətəm] vs “atomic” [ətɒmɪk], which can be viewed as contrastive. On the other hand, others are non-alternating, e.g., “sofa” [səʊfa], or epenthetic, as in the allomorph of the past tense and third singular markers, e.g., “ended” [endɪd]. The latter cases are viewed as non-contrastive examples of schwa vowels. During L2 speech learning, however, learners need to perceptually discriminate an unstressed schwa vowel from stressed full vowels. Therefore, schwa vowels should be perceptually categorized by L2 learners of English as if they have contrast with the other vowels. As mentioned earlier, the PAM assesses the articulatory similarity between native and non-native speech sounds. Based on the assumption that schwa vowels form a contrastive category in L2 learners, the present research investigates the effectiveness of the fundamental and formant frequencies for the prediction of the perceptual assimilation pattern of English schwa vowels by Japanese listeners.

1.3. Preceding Work on Perceptual Assimilation of Schwa Vowels

The perceptual assimilation of English schwa vowels has been reported by Tomaru and Arai [12,13]. They [12] revealed that English schwa vowels were most likely to be perceptually assimilated to the Japanese /a/ at a frequency of 96%. The other 4% were assimilated to /u/ and /o/ vowels (see Appendix A for nonsense words that received an /u/ or /o/ response). Only one /i/ response, accounting for 0.4% of the total responses, was observed. No /e/ response was reported.

However, interestingly, their post-test acoustic analyses could not explain the extreme results of the perceptual assimilation of the schwa by Japanese listeners. Tomaru and Arai [12], suggested that the English schwa vowels used in the experiment might be more like the Japanese /u/ than the Japanese /a/ in terms of the Euclidean distance in the dimensions of first (F1) and second (F2) formant frequencies on the equivalent rectangular bandwidth scale. The following analysis based on the Mahalanobis distance in the F1 and F2 dimensions in Hz using a total of 400 tokens of Japanese /a/ and /u/ vowels from a speaker on the Corpus of Spontaneous Japanese (CSJ) [14,15] found that 76% and 24% of the schwa tokens were categorized as the Japanese /a/ and /u/, respectively [13]. The results of the latter analysis show the tendency that the schwa is more like the Japanese /a/ than the Japanese /u/; however, they are insufficient to explain why schwa vowels are assimilated to Japanese /a/ almost exclusively.

In this study, we attempt to find acoustical cues that Japanese listeners rely on to assess articulatory similarity during the perceptual assimilation of the English schwa to Japanese vowels. Traditionally, in the study of perceptual assimilation, similarity of vowels is evaluated in the F1 and F2 dimensions. However, the results of the former analyses including our own work seem to imply that comparison in terms of F1 and F2 is insufficient to predict the perceptual assimilation pattern. Therefore, the purpose of our research is to probe into the effects of additional information such as the third formant (F3) and fundamental frequencies (F0) on assimilation judgements. Our study consists of two parts. Firstly, we need to confirm that F1 and F2 are insufficient before conducting any analyses involving additional information because the former analysis was conducted without verifying that the stimuli in question were actually schwa vowels. Since schwa tokens were collected through a recording of nonsense words [12,13], it is likely that the collected vowels did not have appropriate schwa quality; this might have caused a misleading outcome of the analysis. Therefore, we first report a preliminary analysis to ensure that the F1 and F2 plane alone cannot explain the perceptual similarities between the English schwa and the Japanese /a/ vowel (Analysis 1). Then, we examine whether additional information such as F3 and F0 helps in predicting the assimilation pattern (Analysis 2).

1.4. Importance of F3 and F0 on Vowel Perception

The addition of F3 and F0 is considered to be effective for the prediction of cross-gender and cross-age vowel normalization [16]. However, this point has never been
discussed in the light of perceptual assimilation. For vowel normalization, Kasuya et al. [16] showed that when there are gender- and age-related changes of formants and pitch, F1 and F2 alone cannot discriminate Japanese vowels because confusion between certain vowels such as /a/ and /o/, and /e/ and /u/ occurs on the F1 and F2 plane. Similar confusion may be involved in the process of perceptual assimilation.

Former studies showed the perceptual assimilation patterns of 11 English vowels, i.e., /i, i, e, et, æ, a, ʌ, ɔ, ou, u, o/ (schwa vowels were not included), by Japanese listeners and some of them analyzed these English vowels in terms of F1 and F2 to explain the assimilation pattern [17, among others]. Although their analysis shows that F1 and F2 information is fairly effective in the prediction of Japanese listeners’ perceptual assimilation, acoustical variation in these dimensions does not give a straightforward explanation for the assimilation patterns. For example, Strange et al. [17] found that the assimilation of the vowel /i/ of a certain speaker could not be accounted for by the formant structure. In addition, although the vowel /ou/ produced by different speakers had very similar formant frequencies, the vowel was assimilated either to the high back /u/ category or to the mid back /o/ category depending on the speaker.

The reason why the former research could not predict the perceptual assimilation patterns of these vowels may be that F1 and F2 alone did not sufficiently reflect articulatory similarities between vowels. For instance, lip rounding, which changes the lip open area and increases the vocal tract length, lowers the frequencies of all formants, i.e., F1, F2 and F3 [18,19]. However, the degree of lowering of each formant depends on the place of constriction in the vocal tract. F3 is associated with palatal constriction, while F2 is associated with velar and velopharyngeal constrictions [19]. F1 is least affected by rounding. Therefore, even though the degree of lip rounding might be reflected in the values of F1 and F2, vowels are well discriminated from each other with additional information of F3 [14,15]. The consonantal and positional contexts of the schwa vowels that were collected and analyzed in previous research [12,13] as well as the five Japanese vowels /i, e, a, o u/. These Japanese vowels were collected from the CSJ [14,15]. The consonantal and positional contexts of the schwa vowels are listed in the Appendix A. The total number of schwa tokens was 55. In order to exclude non-schwa tokens from the analysis, we conducted a screening. Only 17 out of 55 tokens were eligible for the current analysis. The others were excluded from the analysis because (1) they were very weak and had insufficient energy to allow the measurement of the values of formant frequencies or the fundamental frequency and/or (2) they were judged to be non-schwa by a native speaker of English with a phonetic background: all of these were judged to be the Japanese /a/ rather than the English schwa.

2. METHODS

The purpose of the research is to investigate the effects of articulatory similarities that emerge between formant frequencies in the perceptual assimilation of English schwa vowels by Japanese listeners. The first analysis (Analysis 1) attempts to assess the effectiveness of F1 and F2 in the prediction of the assimilation pattern by using a critical band scale, i.e., the Bark scale [23,24]. In the second analysis (Analysis 2), we add F3 and F0 to examine whether the addition of these factors affects the results of a similarity evaluation.

For Analysis 2, we combine the four indices, i.e., F1, F2, F3 and F0, into a single measure following Syrdal and Gopal [25]. According to them, the difference in the formant and fundamental frequency (F0) values on the Bark scale indicates the openness of the mouth and the backness of the tongue: the difference between F1 and F0 values corresponds to the openness of the mouth and the difference between F3 and F2 values corresponds to the backness on the tongue. The critical distance that divides high versus low vowels is 3 Bark: high vowels have an F1 minus (−) F0 difference of less than 3 Bark. In front vowels, the difference between F3 and F2 (F3 − F2) is also less than 3 Bark. In the present study, we call this method of using the distance between formant and fundamental frequencies the Bark difference scale. The scale was proved to be effective in American English vowel discrimination by Syrdal and Gopal [25]. Using the scale, Analysis 2 attempts to examine the effects of the combined information on prediction of the perceptual assimilation pattern.

3. ANALYSIS 1

This section reports the results of a preliminary analysis that investigated the effectiveness of F1 and F2 using the Bark scale.

3.1. Materials

For the present analysis, we used tokens of English schwa vowels that were collected and analyzed in previous research [12,13] as well as the five Japanese vowels /i, e, a, o u/. These Japanese vowels were collected from the CSJ [14,15]. The consonantal and positional contexts of the schwa vowels are listed in the Appendix A. The total number of schwa tokens was 55. In order to exclude non-schwa tokens from the analysis, we conducted a screening. Only 17 out of 55 tokens were eligible for the current analysis. The others were excluded from the analysis because (1) they were very weak and had insufficient energy to allow the measurement of the values of formant frequencies or the fundamental frequency and/or (2) they were judged to be non-schwa by a native speaker of English with a phonetic background: all of these were judged to be the Japanese /a/ rather than the English schwa.
The schwa vowels were spoken by a male American English speaker who was in his forties at the time of the recording. Instructions on how to pronounce unfamiliar nonwords were given by the first author of the former research [12]. For Japanese vowel tokens, we selected recordings of ten male speakers (Appendix B) giving a presentation at an academic conference from the CSJ [14,15]. These speakers were selected because they were reported to have no apparent dialectal accent and a high degree of spontaneity regarding their speech style. From the recorded presentation, we randomly selected 100 samples of each Japanese vowel with clear F0 and formant frequencies. The F0 and formant frequency values were extracted at the vowel’s midpoint using the Praat software [26]. The midpoint of vowels was decided as follows: onset time + ((offset time – onset time)/2). For schwa vowels, the time points of the vowel onset and offset were defined as the time point where the periodic wave of the vowel starts and the time point where it ends, respectively during which the sound has a clear voice bar in its spectrogram. For Japanese vowels from the CSJ, the time points of vowel components that were already labeled to be /i, e, a, o, u/ by the distributors were extracted for the analysis.

3.2. Discrimination Analysis

The purpose of Analysis 1 was to assess the similarities between the schwa and Japanese vowels in the F1 and F2 dimensions using the Bark scale. The Bark scale is employed to compare vowels in terms of perception according to the critical bands. A plot of English schwa and Japanese vowel tokens in the F1 and F2 dimensions on the Bark scale is shown in Fig. 1.

To investigate the similarities between the schwa samples and the five Japanese vowels in the two dimensions, we conducted a discrimination analysis using the Mahalanobis distance. First, we found the Mahalanobis distance between the schwa and each Japanese vowel. Then, we found to which of the Japanese vowels each schwa sample was categorized based on the distance. A schwa was categorized to the Japanese vowel that was minimally distant. The results are summarized in Table 1. The results revealed that 53% (9/17), 41% (7/17) and 6% (1/17) of the schwa tokens were categorized to the Japanese /a/, /o/ and /u/ vowels, respectively. No schwa tokens were categorized to /i/ or /e/. A Kruskal-Wallis test was conducted to confirm that the distances between the schwa and each of the Japanese categories were significantly different from each other. The test found a significant difference overall, $\chi^2(4) = 9.49$, $p < 0.001$. 

![Fig. 1 Plot of F1 and F2 frequencies with 95% confidence ellipses of English schwa vowels (blue) and (a) the Japanese /a/ (red) and /i/ (green), (b) the Japanese /a/ (red) and /e/ (green), (c) the Japanese /a/ (red) and /o/ (green) and (d) the Japanese /a/ (red) and /u/ (green) vowels on the Bark scale.](image)
with a mean rank score of 70.4 for /i/, 66.1 for /e/, 21.9 for /a/, 20.1 for /o/ and 36.6 for /u/. However, a Steel-Dwass test showed no significant difference between the distance between the schwa vowel and the Japanese /i/ vowel and that between the schwa and the /e/ (p > 0.05, Steel-Dwass). Also, the distance between the schwa vowel and the Japanese /a/ and that between the schwa and the /o/ was not significantly different (p > 0.05, Steel-Dwass). Therefore, even though schwa vowels were categorized to the Japanese /a/ more often than to the Japanese /o/, the schwa category was at an equal distance from these Japanese vowel categories.

### 3.3. Interim Discussion

The preliminary analysis revealed that a simple comparison in terms of F1 and F2 could not explain the perceived similarity between the English schwa and the Japanese /a/, mainly because the distance between the schwa and the Japanese vowel /a/ and that between the schwa and the /o/ were statistically the same. In other words, the two Japanese categories are not properly separated using the dimensions of F1 and F2 alone.

One of the main reasons why the discrimination analysis using the F1 and F2 dimensions confuses the /a/ and /o/ vowels may be that these dimensions lack the information associated with lip rounding such as F3. Traditionally, the Japanese /o/ is known as a mid back vowel, which involves relative lowering of the tongue body and raising of the posterior part of the tongue, creating a constriction at the back of the mouth, with a medium degree of mouth opening. The Japanese /a/, on the other hand, is known as a low vowel. This means that the mouth is wide open with the tongue lower than the vowel /o/. However, the place of constriction created by the tongue inside the vocal tract, which is normally referred to as the backness, is not clearly specified for the Japanese /a/: it could be either at the center or back of the mouth [27,28].

In general, the degree of backness is reflected in F2. However, since the Japanese /a/ and /o/ may have the same place of constriction, the difference may not appear in the F2 dimension. Therefore, a clear difference between these vowels should be seen in the degree of lip rounding: the /o/ vowel has less lip open area than the /a/. Since F3 serves as an indicator of the lip open area and the vocal tract length associated with lip rounding, including this factor should make a sufficient distinction between the Japanese /a/ and /o/ vowels. In addition, further information on the degree of mouth opening should help discriminate the two vowels: the mouth opening is greater for /a/ than for /o/. The addition of F0, which is one of the indicators of the openness, should clearly discriminate the vowels from each other. Moreover, in the F1 and F2 dimensions, the Japanese /i/ and /e/ vowels were not clearly separated. This might be because the tongue is relatively high for both vowels, and F1 alone could not show the difference. The addition of F0 and F3 should also explain this point.

### 4. ANALYSIS 2

In Analysis 2, we compare the English schwa and Japanese vowel categories in terms of F1, F2, F3 and F0. In order to express vowel openness and backness in two dimensions by using these measurements, we convert the measurements into the Bark difference scale following Syrdal and Gopal [25]. If Japanese listeners use F3 and F0 as cues in addition to F1 and F2, then a supplement of F3 and F0 information to an analysis should result in distinguishing the Japanese /a/ vowel from the Japanese /o/ vowel and provide supportive evidence that the schwa is more like the Japanese /a/ than the other vowels.

#### 4.1. Materials

The materials used for Analysis 1 were also used for Analysis 2.

#### 4.2. Discrimination Analysis

The Bark difference scale for “openness” is obtained by subtracting the F0 value from the F1 value (F1–F0). Similarly, the value for “backness” is obtained by subtracting the F2 value from the F3 value (F3–F2). The effect of F0 on judgements of openness suddenly changes between F0 = 350 and 400 Hz [22,25], and F0 end correction should be applied in such cases. However, no vowel samples had F0 above 350 Hz, so we report the results based on raw F0 measurements without the end correction.

A plot of the English schwa and the Japanese vowels in terms of the Bark difference scale is shown in Fig. 2. Vowel discrimination analysis was conducted using the same methods as in Analysis 1. The results of the analysis are shown in Table 2. The analysis revealed that most of the schwa samples were categorized to Japanese /a/, i.e., 88% (15/17). Only one stimulus was categorized to /o/.
or to /u/, i.e., 6% (1/17). The Kruskal-Wallis test found a significant overall difference between the average Mahalanobis distance, $\chi^2(4) = 9.49, p < 0.001$, with a mean rank score of 70.4 for /i/, 66.1 for /e/, 21.9 for /a/, 20.1 for /o/ and 36.6 for /u/. A Steel-Dwass test also showed significant difference between the average distance between the schwa and the Japanese vowels except for the distance between the schwa and the Japanese /i/ and that between the schwa and the Japanese /e/ ($p > 0.05$, Steel-Dwass).

The results of Analysis 2 suggest that Japanese listeners use the information of F1, F2, F3 and F0 to assess articulatory similarity during the perceptual assimilation of the schwa. The analysis shows the distinction between the Japanese /a/ and /o/, suggesting that the Japanese /a/ is the closest to the schwa. On the other hand, the analysis did not show the difference in the Japanese /i/ and /e/ vowels. However, since the schwa was hardly ever assimilated to these vowels, it should be safe to conclude that these vowels are sufficiently far apart from the English schwa.

5. DISCUSSION

5.1. Summary and Conclusion

The motivation behind the current research was to find acoustical cues that Japanese listeners rely on to assess articulatory similarity and explain why the English schwa is perceptually assimilated to the Japanese /a/. The purpose of Analysis 1 was to examine whether Japanese listeners rely exclusively on F1 and F2. Analysis 1 compared the English schwa and five Japanese vowel categories, i.e., /i, e, a, o, u/, in F1 and F2 dimensions using the Bark scale. The analysis based on F1 and F2 alone was insufficient to show that the schwa is particularly close to the Japanese /a/. In Analysis 2, we added F3 and F0 information to compare the schwa and the Japanese vowel categories in F1–F0 (openness) and F3–F2 (backness) dimensions using the Bark difference scale. The analyses show that the English schwa is similar to the Japanese /a/ when compared in terms of F1, F2, F3 and F0. These results suggest that, in addition to F1 and F2, Japanese listeners use F3 and F0 as cues to assess
articulatory similarity between the schwa and the Japanese vowels during perceptual assimilation.

5.2. Further Issues

The results of the former perceptual experiment [12] show that 4% of schwa vowels were assimilated to the Japanese /a/ and/or /o/. Eight schwa-containing nonsense words received such responses in addition to the /a/ response. Four of them were included in the current analyses: “azuit” [oːzut], “cabub” [kæbəb], “cadud” [kædəd] and “cazuz” [kæzʊz] (Appendix A). The percentages of /u/ and /o/ responses are as follows: 5% for the /u/ response for “azuit,” 5% for the /o/ response for “cabub,” 10% for the /o/ response for “cadud” and 50% for the /u/ response for “cazuz.” The current analysis may reflect such perceptual results. Analysis 2 revealed that the schwa in “cazuz,” which received a frequent /u/ response, was categorized to the Japanese /u/ vowel when compared in the Bark difference scale.

However, there may be some other factors that affect perceptual assimilation. The current analysis also found that the schwa in “agooke” [oːɡuːk] was categorized to the Japanese /o/ vowel, although it received only the /a/ response in the perceptual experiment. In addition, schwa vowels in the other stimuli, including “azuit,” “cabub” and “cadud,” which received small numbers of /u/ and /o/ responses in the previous perceptual experiment, were all categorized to /a/ in Analysis 2. These results indicate that although the similarities that emerged from the Bark difference scale are mainly responsible for the prediction of perceptual assimilation patterns, there may be some subsidiary cue(s) that is independent of formant energy, such as vowel duration. Further analysis should reveal this point.

In addition, it should be noted that in future research, schwa tokens for perceptual experiments and acoustic analyses should be minimally affected by spelling. The schwa-containing nonsense words analyzed in the current analyses are spelled with an “a” except for the final condition stimuli (Appendix A). The reason we used the same letter for the schwa for almost all nonsense words is that it facilitates pronunciation instructions given to English speakers seeing the nonsense words for the first time. However, it should be noticed that consistent spellings can give a speaker a certain strategy for producing schwa vowels, which may end up adding a constant /a/-like quality to schwa vowels. In fact, several schwas in the medial position were excluded from the current analyses because they sounded like the Japanese /a/ to a native speaker of English; this may be a side effect of the spelling. In order to analyze a large number of samples, schwa production should be carefully controlled in future research.

ACKNOWLEDGMENTS

We would like to express our gratitude to Terri Lander, who gave us comments on the stimuli from a phonetician’s point of view.


This research was supported by JSPS KAKENHI (Grant Numbers 16K16832, 18K12378).

REFERENCES


**APPENDIX A**

The 32 schwa-containing nonsense words analyzed for the current study are listed below according to the position of the schwa-containing syllable. The intended pronunciation is shown in square brackets. See Tomaru and Arai [12] for the full list of the recorded nonsense words used for the perceptual experiments. The words that received an /a/ or /o/ response in addition to an /a/ response are marked with an asterisk. The underlined words were excluded from the current analysis.

(1) Initial
   aguy [ɔːgai], azide [ɔːzaid], aheep [ɔːbip], adeat [ɔːdit], ageek [ɔːgik], aboon [ɔːbuf], adoose [ɔːdus], agooke [ɔːguk], azuit [ɔːzut]*

(2) Medial
   tababite [ˈtɛbəbait], tadalite [ˈtɛdəlait], tagagite [ˈtɛgəɡait], tatazite [ˈtɛtəzait], tabebet [ˈtebəbet], teadadet [ˈtɛdədət], tefagget [ˈtɛfəget], tezzat [ˈtɛzəzet], tebabeat [ˈtiboəbet], teadadet [ˈtidoədit], tegeaget [ˈtiɡəɡit], tezaazet [ˈtɪzəzət], cobaboke [ˈkouboəbouk], codadoko [ˈkoudədouk], cogagoke [ˈkougaəgouk], cozaazoke [ˈkouzaouk], cubabuke [ˈkuboəbuk], cubaduke [ˈkuboəduk], cugaguke [ˈkugoəguk], cuzaazuke [ˈkuzəouk]

(3) Final
   cabub [ˈkæbəb], cadud [ˈkædəd], cazuz [ˈkæzəz]

**APPENDIX B**

Lists of analyzed speakers from the CSJ summarized in Table B·1.

<table>
<thead>
<tr>
<th>Broad Presentation Topic</th>
<th>Speaker ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engineering (Speech Sounds)</td>
<td>A01M0007, A01M0015, A01M0020</td>
</tr>
<tr>
<td>Humanities (Japanese Language)</td>
<td>A02M0076, A02M0098, A02M0107</td>
</tr>
<tr>
<td>Engineering (Natural Language Processing)</td>
<td>A03M0004, A03M0005, A03M0010</td>
</tr>
<tr>
<td>Engineering (Artificial Intelligence)</td>
<td>A04M0026</td>
</tr>
</tbody>
</table>

Kanako Tomaru received her B.A. and M.A. degrees in linguistics from Doshisha Univ., Kyoto in 2009 and Sophia Univ., Tokyo in 2011, respectively. She finished her PhD at the Faculty of Science and Technology, Sophia Univ., in 2015. Her research interests include language perception in relation to the study of Phonetics and Phonology, Acoustics, and Psycholinguistics.

Takayuki Arai received the B.E., M.E. and Ph.D. degrees in electrical engineering from Sophia Univ., Tokyo, Japan, in 1989, 1991 and 1994, respectively. In 1992–1993 and 1995–1996, he was with Oregon Graduate Institute of Science and Technology (Portland, OR, USA). In 1997–1998, he was with International Computer Science Institute (Berkeley, CA, USA). He is currently Professor of the Department of Information and Communication Sciences, Sophia Univ. In 2003–2004, he is a visiting scientist at Massachusetts Institute of Technology (Cambridge, MA, USA). His research interests include signal processing, acoustics, speech and hearing sciences, spoken language processing, and acoustic phonetics.