Production Boundary between Fricative and Affricate in Japanese and Korean Speakers

Kimiko YAMAKAWA1, Shigeaki AMANO2 and Shuichi ITAHASHI1,3

1National Institute of Informatics, Japan
2NTT Communications Science Laboratories, NTT Corporation, Japan
3National Institute of Advanced Industrial Science and Technology, Japan

jin, itabashi@nii.ac.jp, amano@cslab.kecl.ntt.co.jp

Abstract

A fricative [s] and an affricate [ts] pronounced by both native Japanese and Korean speakers were analyzed to clarify the effect of the mother language on speech production. It was revealed that Japanese speakers have a clear individual production boundary between [s] and [ts], and that this boundary corresponds to the production boundary of all Japanese speakers. In contrast, although Korean speakers tend to have a clear individual production boundary, the boundary dose not corresponds to that of Japanese speakers. These facts suggest that Korean speakers tend to have a stable [s]-[ts] production boundary but that differ from Japanese speakers.

Index Terms: fricative, affricate, speech production, non-native speaker, phoneme boundary

1. Introduction

A fricative and an affricate consist of random noise. Previous studies [1, 2, 3] pointed that there are several acoustic features used to distinguish between a fricative and an affricate, such as the duration of a noise, the duration of a rise part of a noise, the duration of a preceding closure, the duration of a preceding vowel, the rate of increase of the sound pressure in the rise part, and the spectral shape of a noise. Most of the previous studies showed these features compared a fricative [ʃ] and an affricate [tʃ]. Their results cannot be directly applied to Japanese fricatives and affricates, because there is a phonemic contrast between a voiceless alveolar fricative [s] and a voiceless alveolar affricate [ts] in Japanese, in addition to a phonemic contrast between [ʃ] and [tʃ]. The [ts] has similar acoustic features to [s]. Both [s] and [ts] consist of a noise but their power patterns are different. Moreover, the [s]-[ts] phonemic contrast is not commonly found in other languages. As a result, it is very difficult for the non-native speakers of Japanese, such as Koreans [4] and Thais [5], to correctly distinguish between [s] and [ts] or to even properly pronounce it because their native languages do not have a [ts].

There are no effective training methods to teach how to correctly pronounce these Japanese consonants for non-native speakers. One of the causes of this problem is that the acoustic features necessary to distinguish between [s] and [ts] are not clearly understood by teachers. If the acoustic features are clarified, it will not only be possible to make scientific methods for Japanese utterance training, but also to develop a Japanese utterance training system using the acoustic engineering technique.

As for the acoustic features of Japanese [s] and [ts], Yamakawa, Amano, and Itahashi [6, 7] modeled the power envelopes of [s] and [ts] as a trapezoidal shape. That is, they divided the power envelopes into the rise part, the steady part, and the decay part. They measured the durations of these parts for [s] and [ts] pronounced by native speakers of Japanese. They found that the production boundary between [s] and [ts] is represented by a linear function with two variables: the duration of the rise part and the sum of the durations of the steady and decay parts (hereafter ”steady+decay”). In other words, the durations of the rise part and steady+decay part are relevant acoustic features to the Japanese [s] and [ts].

This study used these two variables based on their findings, the rise part and steady+decay part durations, to clarify the effect of the mother language on the production of a fricative [s] and an affricate [ts] pronounced by native and non-native Japanese speakers. That is, this study investigated the production boundary between [s] and [ts] of native Japanese speakers and that of native Korean speakers as the non-native speakers. It also investigated relationships between the production boundaries obtained from each speaker to clarify the production consistency among individual native Japanese speakers and among the individual non-native speakers.

2. Experiment

2.1. Participants

Eighteen native Japanese speakers (9 males and 9 females) and 18 native Korean speakers (8 males and 10 females) participated in our experiment. The average age of the native Japanese speakers was 25.3 years (Min = 21, Max = 30, SD = 3.18) and the average age of the native Korean speakers was 25.5 years (Min = 20, Max = 35, SD = 3.94). The native Korean speakers had studied the Japanese language for an average of 20.7 months (Min = 3, Max = 48, SD = 15.08), and they have lived in Japan an average of 10.9 months (Min = 2, Max = 48, SD = 10.64). The participants were paid for their participation.

2.2. Word materials

Eight Japanese words that were in 1-4 mora long with a fricative [s] or an affricate [ts] at the initial phoneme (Table 1) were used as the word materials. The vowel that followed the [s] and [ts] was [i], which was not devoiced. All the word materials consisted of consonant-vowel type morae. Namely, they had no special mora, such as a nasal mora, an obstruct mora, or a lengthened vowel. Two word materials in each word length were configured of a minimal pair. That is, they had the same phoneme sequence except for the initial phoneme, and they had the same accent pattern. There was very little difference in the
auditory word familiarity [8] in each minimal pair. Therefore, the minimal pair probably had no lexical bias in pronunciation.

2.3. Procedure

The recording of the word materials was conducted in a quiet room. The pronunciation list of all the word materials were duplicated 4 times and randomized for each participant. Therefore, there were a total of 32 words on the list. In each trial, one of the words on the list was presented on a computer screen in Japanese hiragana orthography. The participant was asked to push the start button and then naturally pronounce the presented word at a normal speaking rate.

The participant’s pronunciation was recorded using a microphone (ATM31a, AudioTechnica) and a A/D converter (UA25-EX, Roland) with 16-bit quantization and 48-kHz sampling frequency, and then was stored in a computer.

When the participant finished each pronunciation, she/he was asked to push the stop button. The computer automatically checked the recorded pronunciation. It gave an alert if the loudness of the pronounced word was too low or too high, or if the beginning or end of the pronounced word was not properly recorded. In these cases, the word was recorded again at that time. In addition to the computer checking, an operator monitored the pronunciation, and if problems, such as mispronunciation or hesitant pronunciation were found, the words were rerecorded at the end of the experiment.

For each word, the root mean square power was calculated with a 6 ms window size and a 1 ms window shift. One author measured the start point of the rise part (a), the start point of the steady part (b), the start point of the decay part (c) and the end point of the decay part (d) for [s] and [ts] in milliseconds by looking at the waveform and the root mean square power pattern of each word (Fig. 1). Another author checked and corrected the measured points using the same procedure. The durations of the rise and steady-decay parts were calculated from these measured points.

2.4. Results

2.4.1. Native Japanese speakers

The average and standard deviation of the rise part and a steady-decay durations are listed in Table 2. The durations of these parts of [s] tend to be longer than those of [ts].

When a discriminant analysis was performed for an individual native Japanese speaker, the averaged discriminant error rate was 1.22% (SD = 2.18). This low error rate indicates that each native Japanese speaker distinguishes between [s] and [ts] in their pronunciation with a very clear criterion.

When a discriminant error was calculated for an individual native Japanese speaker using the production boundary obtained from all the native Japanese speakers (Eq. 1), the averaged discriminant error rate was 3.82%. This very small discriminant error indicates that the productions of [s] and [ts] are very well discriminated by this production boundary. In other words, the Japanese production boundary between [s] and [ts] is clearly represented using a linear function with the variables of the rise part and steady-decay durations, which is consistent with the results from previous studies [6, 7].

When a discriminant analysis was performed for an individual native Japanese speaker, the averaged discriminant error rate was 3.82% (SD = 5.63). This low error rate indicates that Eq. 1 does quite well at discriminating [s] and [ts] pronounced by each native Japanese speaker.

2.4.2. Native Korean speakers

The average and standard deviation of the rise part and steady-decay part durations are listed in Table 3. The durations of these parts of [s] tend to be longer than those of [ts].

### Table 1: Word materials.

<table>
<thead>
<tr>
<th>Word</th>
<th>Length (mora)</th>
<th>Accent pattern</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>/su/</td>
<td>1</td>
<td>H</td>
<td>vinegar</td>
</tr>
<tr>
<td>/tsu/</td>
<td>1</td>
<td>H</td>
<td>harbor</td>
</tr>
<tr>
<td>/sura/</td>
<td>2</td>
<td>LH</td>
<td>do</td>
</tr>
<tr>
<td>/tsura/</td>
<td>2</td>
<td>LH</td>
<td>fish</td>
</tr>
<tr>
<td>/sune/</td>
<td>3</td>
<td>LHL</td>
<td>silk</td>
</tr>
<tr>
<td>/tsune/</td>
<td>3</td>
<td>LHL</td>
<td>pinch</td>
</tr>
<tr>
<td>/sumago/</td>
<td>4</td>
<td>LHLL</td>
<td>single-string harp</td>
</tr>
<tr>
<td>/tsumago/</td>
<td>4</td>
<td>LHLL</td>
<td>multi-string harp</td>
</tr>
</tbody>
</table>

### Table 2: Mean duration (ms) of rise and steady-decay parts of [s] and [ts] pronounced by native Japanese speakers. The standard deviation is shown in parentheses.

<table>
<thead>
<tr>
<th>Length (mora)</th>
<th>Phoneme</th>
<th>Rise part</th>
<th>Steady+Decay part</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[s]</td>
<td>109.1(29.0)</td>
<td>74.7(25.4)</td>
</tr>
<tr>
<td>2</td>
<td>[ts]</td>
<td>32.7(19.0)</td>
<td>62.5(22.1)</td>
</tr>
<tr>
<td>3</td>
<td>[s]</td>
<td>90.0(27.5)</td>
<td>76.6(31.6)</td>
</tr>
<tr>
<td>4</td>
<td>[ts]</td>
<td>35.5(18.7)</td>
<td>63.5(21.5)</td>
</tr>
<tr>
<td>1-4</td>
<td>[s]</td>
<td>94.5(30.2)</td>
<td>77.1(29.9)</td>
</tr>
<tr>
<td></td>
<td>[ts]</td>
<td>34.6(18.4)</td>
<td>64.2(21.5)</td>
</tr>
</tbody>
</table>

Figure 1: Diagram of power envelope of consonant [s] and [ts].

$$ y = -1.478x + 166.11 $$ (1) 

where $x$ is the rise part duration and $y$ is the steady-decay part duration. In Fig. 2, this production boundary is shown with a solid line.

The discriminant error rate with this production boundary was 3.82%. This very small discriminant error indicates that the productions of [s] and [ts] are very well discriminated by this production boundary. In other words, the Japanese production boundary between [s] and [ts] is clearly represented using a linear function with the variables of the rise part and steady-decay durations, which is consistent with the results from previous studies [6, 7].
Figure 2: Scatter plot of production data of [s] and [ts] by native Japanese speakers with variables of rise part and steady+decay part durations. The solid line is the production boundary between [s] and [ts] (Eq. 1).

Figure 3 is a scatter plot of [s] and [ts] with the variables of the rise part duration and steady+decay part durations. A discriminant analysis was performed with the variables of the rise and steady+decay durations to obtain the production boundary between [s] and [ts] for native Korean speakers. The obtained production boundary was

\[ y = -2.103x + 194.18 \]  

where \( x \) is the rise part duration and \( y \) is the steady+decay part duration. In Fig. 3, the solid line is the production boundary between [s] and [ts] for native Korean speakers.

The discriminant error rate with this production boundary was 13.89%. This discriminant error rate is not small, hence the productions of [s] and [ts] are not well discriminated by this production boundary.

The dashed line in Fig. 3 represents the production boundary by the native Japanese speakers (Eq. 1). This dashed line is different from the solid line representing the production boundary of the native Korean speakers (Eq. 2). The discriminant error rate of Korean’s [s]-[ts] production with the Japanese production boundary was 15.63%. These indicate that the native Korean speakers have different production boundaries than the native Japanese speakers.

When the discriminant analysis was performed for each individual native Korean speaker, the averaged discriminant error rate was 5.90% (SD = 11.89). This low error rate indicates that each native Korean speaker pronounces [s] and [ts] with a clear discrimination according to her/his own criterion.

When a discriminant error was calculated for an individual native Korean speaker using the production boundary obtained from all the native Japanese speakers (Eq. 1), the averaged discriminant error rate was 15.63% (SD = 17.14). This high error rate indicates that native Korean speakers have different production boundaries than native Japanese speakers. The large standard deviation indicates that some of the native Korean speakers have very different production boundaries than native Japanese speakers.

Table 3: Mean duration (ms) of rise part and steady+decay part of [s] and [ts] pronounced by native Korean speakers. The standard deviation is shown in parentheses.

<table>
<thead>
<tr>
<th>Length (mora)</th>
<th>Phoneme</th>
<th>Rise part</th>
<th>Steady+Decay part</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[s]</td>
<td>108.9(37.6)</td>
<td>73.1(37.6)</td>
</tr>
<tr>
<td></td>
<td>[ts]</td>
<td>35.2(33.2)</td>
<td>51.1(30.8)</td>
</tr>
<tr>
<td>2</td>
<td>[s]</td>
<td>104.0(33.1)</td>
<td>62.5(29.1)</td>
</tr>
<tr>
<td></td>
<td>[ts]</td>
<td>33.8(31.9)</td>
<td>49.3(24.0)</td>
</tr>
<tr>
<td>3</td>
<td>[s]</td>
<td>88.7(32.4)</td>
<td>65.7(23.1)</td>
</tr>
<tr>
<td></td>
<td>[ts]</td>
<td>29.4(23.1)</td>
<td>50.0(27.4)</td>
</tr>
<tr>
<td>4</td>
<td>[s]</td>
<td>84.8(29.8)</td>
<td>62.6(25.3)</td>
</tr>
<tr>
<td></td>
<td>[ts]</td>
<td>32.0(26.3)</td>
<td>52.3(25.1)</td>
</tr>
</tbody>
</table>

3. Discussion

As shown in Sec. 2.4.1, native Japanese speakers had small error rates for the production boundary from all the Japanese speakers and for an individual production boundary. The relationships of these two error rates are shown in Fig. 4. It means that each native Japanese speaker has a clear production boundary between [s] and [ts], and that the production boundary of each native Japanese speakers coincides with the production boundary of all native Japanese speakers. In other words, each native Japanese speaker has the same production boundary between [s] and [ts].

By contrast, native Korean speakers had large error rates when using a Japanese boundary and small error rates when using an individual Korean boundary, as discussed in Sec. 2.4.2, except that one participant had a large error rate when using an individual Korean boundary (Fig. 4). This means that each native Korean speaker has a rather clear production boundary between [s] and [ts], but the boundary does not coincide with the production boundary of native Japanese speakers.
Table 4: Error rate (%) in [s]-[ts] discrimination using Japanese production boundary.

<table>
<thead>
<tr>
<th>Length (mora)</th>
<th>Japanese speaker</th>
<th>Korean speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>6.90</td>
</tr>
<tr>
<td>2</td>
<td>1.74</td>
<td>7.29</td>
</tr>
<tr>
<td>3</td>
<td>3.47</td>
<td>7.64</td>
</tr>
<tr>
<td>4</td>
<td>2.43</td>
<td>9.38</td>
</tr>
</tbody>
</table>

Table 4 lists the error rates of [s]-[ts] discrimination for each word length by using the production boundary from the all Japanese speakers. The result of a variance analysis on the error rate listed in Table 4 revealed that the difference between native Japanese and Korean speakers was significant [$F_{1,3} = 78.81$, $p<.01$], but the difference in the word lengths was not significant. Therefore, the error rates of the [s]-[ts] discrimination when using a Japanese boundary is larger for Korean speakers than Japanese speakers irrespective of the word length.

It is possible that native Korean speakers’ error rates of [s]-[ts] discrimination when using a Japanese boundary would decrease if they had a longer education in the Japanese language or if they stayed longer in Japan. However, there were no clear relationships between the error rate and the period of studying Japanese or period of living in Japan. It might be that these variables concerning the length of exposure to Japanese language were not relevant to the pronunciation ability of [s] and [ts]. Other variables such as the perceptual ability to distinguish the phonemes of foreign languages might affect pronunciation ability.

No clear relationships between the error rates and lengths of exposure to Japanese language suggest that native Korean speakers with more exposure to Japan do not necessarily have a better pronunciation ability for discriminating between [s] and [ts]. The consciousness-raising [9] might be necessary to improve the pronunciation ability. The duration measurement used in this study would provide an objective measure of the pronunciation ability and would help raise the consciousness for non-native speakers.

Words with [s] and [ts] only in the initial phoneme was used in this study. Words with [s] and [ts] in the medial or final phoneme should be used to check whether the current results can be dilated to other phoneme positions.

The word materials in this study were pronounced at a normal speaking rate rather than at a fast or slow speaking rate. Previous studies showed that speaking rate affects production boundaries (e.g., [10]). Therefore, it is possible that the speaking rate might affect the production boundary between [s] and [ts] for both native Japanese and Korean speakers. The effect of the speaking rate should be revealed in a future study.

4. Conclusion
A production analysis of a fricative [s] and an affricate [ts] of Japanese for native Japanese and Korean speakers revealed that native Japanese speakers have a clear individual production boundary between [s] and [ts], and that the individual production boundary corresponds to the production boundary obtained from native Japanese speakers. In contrast, although native Korean speaker tend to have a clear individual production boundary between [s] and [ts], the individual production boundary does not correspond to the production boundary of native Japanese speakers. These facts suggest that native Ko-