Perceptual Space of English Fricatives for Japanese Learners

Won Tokuma\(^1\) and Shinichi Tokuma\(^2\)

\(^1\)Seijo University, Tokyo, Japan  
\(^2\)Chuo University, Tokyo, Japan  
w-tokuma@mvg.biglobe.ne.jp

Abstract
This study investigates the relationship between perceptual and physical spatial configurations of English fricatives judged by Japanese learners. Multidimensional scaling analyses were used to obtain spatial representations from similarity judgements and spectral distance analysis. The 2-dimensional perceptual solution did not correspond well to its physical solution. For the physical space, the two dimensions could be interpreted as the phonetic feature properties of ‘sibilance’ and ‘place’ of articulation for fricatives. For the perceptual space, the sibilance property was maintained while the place property was not. In terms of their spectral properties, overall spectral shape was a salient acoustic cue for L2 learners but individual spectral peak frequencies seem to be language specific. The results imply that L2 perception was not based on the acoustic signals, but largely influenced by the learners’ native language phonological structure.

1. Introduction
The common practice in traditional second-language perception experiments, is to test a hypothesis about the perceptual significance of certain acoustic characteristics of signals. However, it has been difficult to assign a single coherent set of acoustic measures to describe all of the consonants. This is because consonants differ significantly among themselves in their acoustic properties. Because of these differences, the perception of L2 consonants cannot be modelled as a whole in terms of a small universal acoustic parameter, like duration or formant patterns in the case of vowel perception. Instead, it is usually necessary to be studied in the subsets of the whole consonant inventory, and predominantly in minimal pair contrasts, such as /f/-/θ/.

Instead, we concentrate on similarity data, which can be used to build spatial models of each level of perceptual processing - acoustic signals, perceptual judgments, and phonetic contrasts. At each level of representation, a spatial map is used to indicate the relative location of units on measurement dimensions where distance is inversely related to the similarity. The phonetic units are recognised in terms of regions they occupy in such spaces with respect to the other units.

This approach is consistent with models of speech perception such as the prototype theory [1], according to which the correct identification of speech segments depend on the perceived distances between speech stimuli and a prototype/region in perceptual space.

Most previous works based on spatial representations were on vowels [2] and they have revealed a high correspondence between acoustic differences and spatial distances based on perceptual similarity measures. Perceptual dimensions have correlated highly with acoustic measurements of vowels, F1, F2 and F3, so vowel perceptual spaces resemble traditional acoustic maps of vowel inventories.

This spatial approach was also used to investigate a set of English consonants [3]. The study enabled us to identify key factors involved in each domain of the processing, and to demonstrate simple correlation across the different domains. The auditory and perceptual dimensions have clear phonetic interpretations, and are also related to concrete physical properties of fricative spectra. This is a novel finding, to the extent that spatial representations had never before been clearly established for consonants; as a consequence, the relationship between the spatial representations across the different domains had never been open to investigation.

The weakness of an approach only based on a small set of consonants is that it may miss the simple and robust features of sounds that characterise the second-language processing as a whole. This spatial representation approach, initially developed for the analysis of vowels, could provide an alternative to a conventional approach to L2 consonant perception.

The following section describes the results of the two experiments on native-language perception for English fricatives. Section 3 is about the spatial representation of L2 perception. Finally section 4 discusses the relationship between perceptual physical and phonetic domains for L2 speech processing.

2. Perceptual and Physical analyses for L1

2.1. Materials
The previous work on the spatial representations of consonants used voiceless English fricatives, /θ s f j h/, followed by the vowel /a/. To avoid the criticism that the data were limited to male voices, the physical representations were obtained from male speakers, while the perceptual test was based on a female voice of the same fricative material. All the recordings were made in an anechoic room onto a Sony DTC-1000ES digital audio tape recorder. They were digitized with a 20kHz sampling frequency and 16-bit quantization, and transferred onto a computer disk.

2.2. Physical space
The physical space was obtained in four main stages. Firstly, the spectra were processed by a simple 1/3-octave bandpass filtering to model filter bank analyses in the auditory periphery. The intensity axis was also transformed into a logarithmic scale, to reflect the non-linear loudness density pattern in the auditory periphery. The outcome is an auditory excitation pattern. 32-channel filters were used for the Euclidean metric analysis. To account for the dynamic fluctuation of the fricative signal, and differences in the length between the different fricatives and speakers, a non-linear time alignment technique was used [4]. This technique is
Based on a principle of optimization, which relies on finding the shortest path between two compared segments aligned on a graph. Finally, the distances between the aligned auditory spectra were calculated for each production of each speaker and used for 3-way, triangular matrices, interval level MDS analysis. The object of this technique is to obtain an optimal spatial representation of the scaled objects on the basis of analysed distances. In this way, we determine the minimal number of physical dimensions required to model the production data with maximal variance in the data accounted for. Thus the dimensions and axes are not predetermined.

2.2. Perceptual space

As mentioned before, the stimuli for the perception test were the same fricatives but read by a female native speaker of R.P. After the digitization, the fricatives were excised and normalized in their intensity with respect to the RMS levels.

20 students were paid to listen to stimuli pairs of AB AC and each time had to choose the more similar sounding pair. A perceptual similarity matrix was constructed from the responses of each subject. A weighted MDS analysis with the square matrix option was carried out on the similarity matrices. The badness-of-fit curve and the interpretability of spatial arrangements suggest that a 2-dimensional solution is most appropriate to model the data. Fricatives are clearly organized in terms of their ‘place’ and ‘sibilance’.

2.3. Spectral analyses

The average spectral shape was obtained in three separate stages. First, the output energy levels of each auditory filter were averaged across the whole length of each fricative segment. In this way, for every individual production, a series of 32 numbers was obtained, representing 32 filter bands. In order to accommodate the differences in the overall level of the fricative segments, the output levels of the 32 bands were reduced by the mean level of that particular production. This process was repeated for each production of each fricative.
spectra are mainly flat. /s/ and /ʃ/ can be characterised by a single broad-band peak; however, the low cut-off frequency occurs a little higher for /s/ at around 3600 Hz, than /ʃ/, at 2000 Hz. For /h/, the spectral peaks occur at around 770 Hz and 2000 Hz, which correspond to the formant frequencies of the following [ə] vowel.

Dimensions in MDS spaces have corresponded quite closely to spectral properties of ‘peakiness’ – the maximum distance to mean amplitude – and the centre of gravity of the average spectra.

In summary, it was shown that there is a close correlation between physical and perceptual configurations for English voiceless fricatives. The configurations were obtained from multidimensional scaling of similarity judgements and spectral distance analyses. These could also be easily associated with their phonetic feature properties of ‘place’ and ‘sibilance’, as well as the spectral properties of ‘centre of gravity’ and ‘peak frequency’. Thus, it can be concluded that L1 perception of fricatives is mainly based on their physical properties, as the perceptual mechanism of vowels is represented by the vowel quadrilateral [5].

3. L2 perception

3.1. Procedure
For cross-language comparison, 19 Japanese learners of English listened to the same English fricative stimuli used in the L1 perceptual experiment (Section 2.2). They were first-year and second-year students of Kunitachi College of Music, in Tokyo. They studied English for 6 years at secondary and high schools, and none of them had lived in an English speaking country. They all took the author’s English class, and their English command is at the false-beginner or pre-intermediate level. The experiment was conducted in a quiet sound-proof classroom in the College. The stimuli were played through speakers equipped in the classroom.

Perceptual data were accumulated over trials by assigning 1 scores for the fricative pairs selected as more similar, and 0 scores for the pairs which were not selected. An example of a similarity matrix for the five fricatives is given below.

<table>
<thead>
<tr>
<th></th>
<th>f</th>
<th>θ</th>
<th>s</th>
<th>ʃ</th>
<th>h</th>
</tr>
</thead>
<tbody>
<tr>
<td>f</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>θ</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>s</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>ʃ</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>h</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Similarity matrix for a Japanese learner who rated the English fricative pairs.

The similarity matrices obtained from each of the listeners were typically not symmetrical. Thus, the square matrix option was used in MDS analysis (ALSCAL program, SPSS Windows version 11.0). The blank entries were given an appropriate large number for the distance conversion.

3.2. Results and discussion
The badness-of-fit curve and the interpretability of spatial arrangements suggest that a 2-dimensional solution is most appropriate to model the data. This perceptual space is presented in Figure 4.

![Figure 4](image-url)

Figure 4. The perceptual space of English voiceless fricatives for 19 Japanese learners.

Each fricative area is identifiable but we can observe considerable overlaps in the perceptual map. The ‘sibilance’ property was maintained but the ‘place’ property is not. /ʃ/ and /h/ are closer than /f/ and /θ/. This can be explained by their neutralised position before the vowel /u/ in Japanese word like *fujisan* ‘Mt. Fuji’. /h/ is pronounced as a bilabial fricative, /ɸ/, in English. The Japanese bilabial fricative is somewhat like a cross between the English phoneme /θ/ and /h/. Japanese learners sometimes mispronounce English words like ‘who’.

/s/ and /ʃ/ are relatively distinct. It is interesting to compare their distribution with that of the perception of Japanese fricatives, /ʃ/, /f/ and /θ/. Figure 5 is the perceptual map of Japanese fricatives judged by Japanese listeners [7]. Each fricative area is smaller but less distinct. This reflects their neutralised pronunciations. /ʃi/ in English becomes /ʃi/ in Japanese loan words like /ʃisoe/ ‘sesso’ and /ʃizun/ ‘season’, /ʃ/ and /θ/ also tend to merge particularly for Tokyo dialect speakers. They pronounce /ʃidoi/ ‘terrible’ as /ʃidoi/ [6]. The subjects in this experiment were all standard Japanese speakers, but mainly from the Greater Tokyo area. These neutralised pronunciations between /ʃi/, /ʃu/ and /ʃi/ in Tokyo dialect may also affect the perceptions of these fricatives before other vowels. Therefore, English /s/ and /ʃ/ seem to cause less confusion for Tokyo dialect speakers.

/θ/ and /ʃ/ are quite distinct because /ʃ/ is not in Japanese phoneme inventory. /θ/ in English loan word ‘thank-you’ is often pronounced as ‘sank-you’, so the perceptual configuration in the map reflects this characteristic Japanese pronunciation.
4. Conclusions

The above results indicate that for L2 perception, the occupation of the space differs and the imputed criteria for identifying a segment, change in comparison to L1 perception. The major difference between L1 and L2 perceptions was that for L1 perception, the perceptual configurations could be driven from the physical configurations, while for L2 perception, the correlation between the two spaces was low. This shows that the perception of non-native listeners is not predictable from the acoustic signals; instead the perceptual map was influenced by their L1 phonological structure.

The ‘sibilance’ property was salient for both English native speakers and Japanese learners, while the ‘place’ distinction was obscured for the L2 perception. In terms of the spectral analyses, this result implies that the overall spectral shape property is a universal perceptual cue, whereas the spectral peak frequency may be language specific.

This spatial study supports second language perception models such as the Native Language Magnet Model [8]. According to this model, the easiest sounds to acquire in a second language are those in which the prototypes are almost identical in both languages. Each prototype is the centre of perceptual area for each sound in the map. L2 sounds that are outside the L1 acoustic space are somewhat harder or slower to learn, but even harder are those sounds that are non-prototypical instances of an L1 sound. Therefore, English /s/ and /ʃ/ before the vowel /a/ almost coincide with Japanese /s/ and /ʃ/, while English /θ/ does not exist in Japanese. But English /θ/ seems to be the non-prototypical instances of Japanese /ɕ/, resulting in the overlap in the perceptual map with /h/.

However, the results obtained from the experiment is far from conclusive. It is yet to be investigated whether or how the English command of listeners, affects the perceptual map of English fricatives, since the listeners’ English command clearly influences the perceptual score. Future work includes testing Japanese learners at different English levels (false-beginner, intermediate and advanced) and describing the perceptual difference between them.

5. Acknowledgements

We wish to thank Mark Huckvale for his continuing support and encouragement. We are also grateful to Stuart Rosen and other members of Wolfson House, University College London, for their help and advice. This research was partially funded by a Chuo University Grant for Special Purposes.

6. References