SPECTRAL MEASURES FOR SIBILANT FRICATIVES OF ENGLISH, JAPANESE, AND MANDARIN CHINESE

Fangfang Li[a], *Jan Edwards*[b], *& Mary Beckman*[a]

[a]Department of Linguistics, OSU & [b]Department of Communicative Disorders, UW
{cathy|mbeckman}@ling.ohio-state.edu, jedwards2@wisc.edu

ABSTRACT

Most acoustic studies of sibilant fricatives focus on languages that have a place distinction like the English distinction between coronal alveolar /s/ and coronal post-alveolar /ʃ/. Much less attention has been paid to languages such as Japanese, where the contrast involves tongue posture as much as position. That is, the Japanese sibilant that contrasts with /s/ is /ɕ/, an alveolopalatal fricative that has a “palatalized” tongue shape (a bunched predorsum). This paper describes measures that can be calculated from the fricative interval alone, which we applied both to the place distinction of English and the “palatalization” or posture distinction of Japanese. The measures were further tested on Mandarin Chinese, a language that has a three-way contrast in sibilant fricatives contrasting in both tongue position and posture.

Keywords: Spectral analysis, Sibilant fricatives

1. INTRODUCTION

Although both English and Japanese have a 2-way contrast in sibilant fricatives, they contrast in different articulatory aspects. In English, dental or alveolar /s/ contrasts with coronal post-alveolar /ʃ/ in tongue position, with the narrowest constriction for /ʃ/ being further back in the oral cavity than that for /s/, whereas in Japanese, dental /s/ contrasts with alveolopalatal /ɕ/ primarily in tongue posture, with the front of the tongue body being bunched up towards the palate in producing /ɕ/, but not in the production of /s/ [7, 10, 18]. Mandarin Chinese is a language that has a 3-way contrast among dental/alveolar /s/, alveolopalatal /ɕ/, and retroflex postalveolar /ʂ/, where /ɕ/ contrasts with /s/ and /ʂ/ in posture, and /ʂ/ contrasts with /s/ and /ɕ/ in place [11, 18].

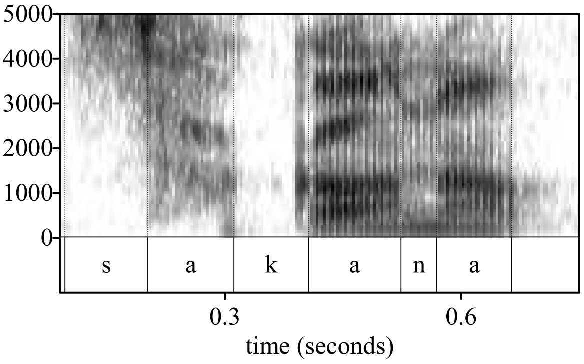
Extensive research has been done on languages that have a place distinction in sibilant fricatives, especially English. It has been suggested that English /s/ and /ʃ/ can be differentiated by the spectral properties of the frication itself [1, 8]. As mentioned earlier, English /s/ and /ʃ/ contrast in place, with the narrowest lingual constriction being made more backward in the oral cavity for /ʃ/ than for /s/. The longer front cavity in producing /ʃ/ lowers the overall frequency for the major energy concentration in the fricative spectrum. The length difference is further enhanced by lip protrusion in /ʃ/, so that it has been consistently observed that there is more low-frequency energy for the /ʃ/ spectrum and more high-frequency energy for /s/ [8, 16].

This generalization about a difference in energy distribution between /s/ and /ʃ/ can be captured effectively by the centroid frequency, the first moment or mean frequency when the power spectrum is treated as a probability distribution [3, 9, 15]. This is a measure that negatively correlates with the length of the front resonating cavity, and thus roughly describes where the constriction is made relative to the length of the oral cavity.

This interpretation of the centroid as primarily reflecting front-cavity resonances fails in fricatives such as the Mandarin retroflex /ʂ/. The short slack constriction used in producing /ʂ/ allows coupling of the front cavity with the back cavity and therefore creates a low-frequency prominence in the fricative spectrum [17]. If the centroid is to be calculated over the whole spectrum, the low centroid value reflects the back cavity resonances as much as it does the front cavity resonances. Therefore, this measure needs to be refined in order to assess front cavity resonances in the sibilant fricatives of Mandarin Chinese.

To assess a tongue posture distinction, such as that in Japanese, Polish, or Mandarin Chinese, several researchers [4, 7] have proposed using the F2 frequency taken at the onset of the vowel following the fricative to index the length of the back cavity in the fricative. A “palatalized posture” in producing alveolopalatal fricative /ɕ/ creates a long palatal channel and a shorter back cavity, which yields higher onset F2 frequency than otherwise. However, this onset F2 frequency measure varies much more with the following vowel than do measures taken from the fricative interval, which makes it less appropriate for cross-linguistic comparison. Also, it cannot be a reliable measure for the Tokyo dialect of Japanese, where vowels are frequently devoiced or even deleted [13], as shown in Fig. 1. Therefore, in this paper, we explore better measures that can be calculated from the fricative interval alone, to see whether any could be applied both to the place contrast between the two coronal fricatives of English and of Mandarin, and to the posture contrast between coronal fricatives and the alveolopalatal fricative of Japanese and Mandarin.

Figure 1: Production of *sakana* 'fish' by a male adult speaker of Tokyo Japanese, with the first [a] devoiced.



2. METHODS

2.1. Materials and Data collection

The materials were lists of words beginning with the target consonants followed by one of the five vowels /a/, /i/, /o/, /e/ or /u/, if allowed by the phonotactics of the language. For example, Japanese /s/ was not elicited in the context of /i/, since it is phonotactically illegal. Also, the English /i/, /e/, and /u/ contexts included both the tense and the lax vowels. There were three word targets for each vowel context. Thirty speakers aged from 18 to 30 years were recruited for this study, with all U. S. English speakers recorded in Columbus, Ohio, all Japanese speakers in Tokyo, and all Mandarin speakers in Songyuan, China, to ensure dialect homogeneity. For each language, there were 5 males and 5 females. Productions were elicited in a word-repetition task, with each word prompted by a digitized audio stimulus played through good-quality loud-speakers connected to the audio port of an IBM Thinkpad notebook computer, along with a picture prompt presented on the computer screen. All productions were recorded using a Marantz 660 Flashcard recorder.

2.2. Developing Spectral Measures

Halle & Stevens [7] examine acoustic properties of the Polish alveolopalatal fricative /ɕ/ and retroflex fricative /ʂ/, and point out that the two fricatives differ in the presence or absence of a spectral peak in the F2 region of the fricative noise spectrum. This is because the "palatalized" posture in producing /ɕ/ forms a long palatal channel, which effectively prevents acoustic coupling to the back cavity, and results in little resonance in the F2 region of the fricative spectrum. By contrast, the short constriction made by the retroflex fricative /ʂ/ yields a spectral peak in the F2 region owing to the considerable acoustic coupling with the back cavity. Halle and Stevens locate the F2 region at around 1500 Hz, the corresponding F2 of a neutral vowel as produced by typical American white male speakers [2]. Their results agree with those of Fujisaki & Kunisaki [5], who model the acoustics of voiceless fricatives in Japanese, and show that adding a zero in the model for /ɕ/ at a point around 1500 Hz effectively differentiates /ɕ/ from /s/.

In this study, two spectral parameters are proposed. Both measures are taken from Bark spectra calculated over 20 ms windows from the middle of the fricative noise interval. The first is the Amplitude Ratio ("ampRatio" hereafter), which is the difference in dB between the amplitude of the most prominent peak and the F2 amplitude. This measure is designed to assess the degree of "palatalization" — i.e., the tongue posture difference. For alveolopalatal /ɕ/, the value of ampRatio is predicted to be larger due to the lack of resonance in the F2 region, whereas that of /ʂ/ is expected to be much smaller due to the coupling between the front and the back cavities.

This "ampRatio" is similar to the relative amplitude measure developed by McGowan & Nittrouer [14] for English children's productions. However, our "ampRatio" differs from their relative amplitude measure in how we identify the F2 region. McGowan and Nittrouer measured an observable F2 peak in the spectrum, with the F2 peak being usually prominent in children's speech.

In this paper, since what we look for is the amplitude of an F2 anti-formant, it is hard to locate it directly from the spectrum. Instead, the frequency region of F2 was calculated by taking the 3/5 ratio of the F3 frequency of the back vowel /a/ for each speaker. This approximates the F2 for a theoretical neutral vowel, which Japanese and Mandarin Chinese do not have.

The second measure is a refinement of a standard measure for the place distinction. It is the centroid frequency calculated for the frequency band above this "F2 region" (CentHigh hereafter). Since spectral energy in the low frequency range can reflect a coupled back cavity resonance, calculating the centroid only in the high frequency band makes it a more precise measure of front cavity resonance. More specifically, for example, the F2 region for the Japanese female speakers is in the 12th Bark, and CentHigh was calculated over the range from 12th Bark to 29th Bark.

3. RESULTS

Fig. 2 plots the ampRatio against the CentHigh for all fricatives produced by the five female speakers in all vowel contexts for each language. The measures for the male speakers patterned similarly.

It is clear from the first panel in the figure that English /s/ and /ʃ/ can be differentiated by the parameter CentHigh, with /s/ (circles) having higher energy concentration than /ʃ/ (triangles). This is in accordance with the articulatory difference, since /s/ is produced with a more anterior constriction than /ʃ/, as well as with no lip protrusion.

For the Japanese results in the second panel, on the other hand, the contrast between /s/ (circles) and /ɕ/ (pluses) relies more on the AmpRatio, although both parameters seem to contribute to the distinction jointly. A linear discriminant analysis (LDA) using both parameters identified the category of 92% of the tokens. The separation by ampRatio is consistent with the predictions from what is known of the articulation, with a higher value for /ɕ/ than for /s/, suggesting a more "palatalized" (bunched predorsum) posture for /ɕ/.

In the case of Mandarin sibilants in the third panel, retroflex /ʂ/ (triangles) is separated from the other two fricatives in the CentHigh dimension as well as in ampRatio, suggesting that the position of the narrowest lingual constriction is further back in the oral cavity, and also that the tongue shape is apical, which allows much coupling with the back cavity. At the same time, /s/ and /ɕ/ are separated jointly by the two parameters, mimicking the /s/-/ɕ/ contrast pattern in Japanese. The LDA again correctly identified 92% of the tokens.

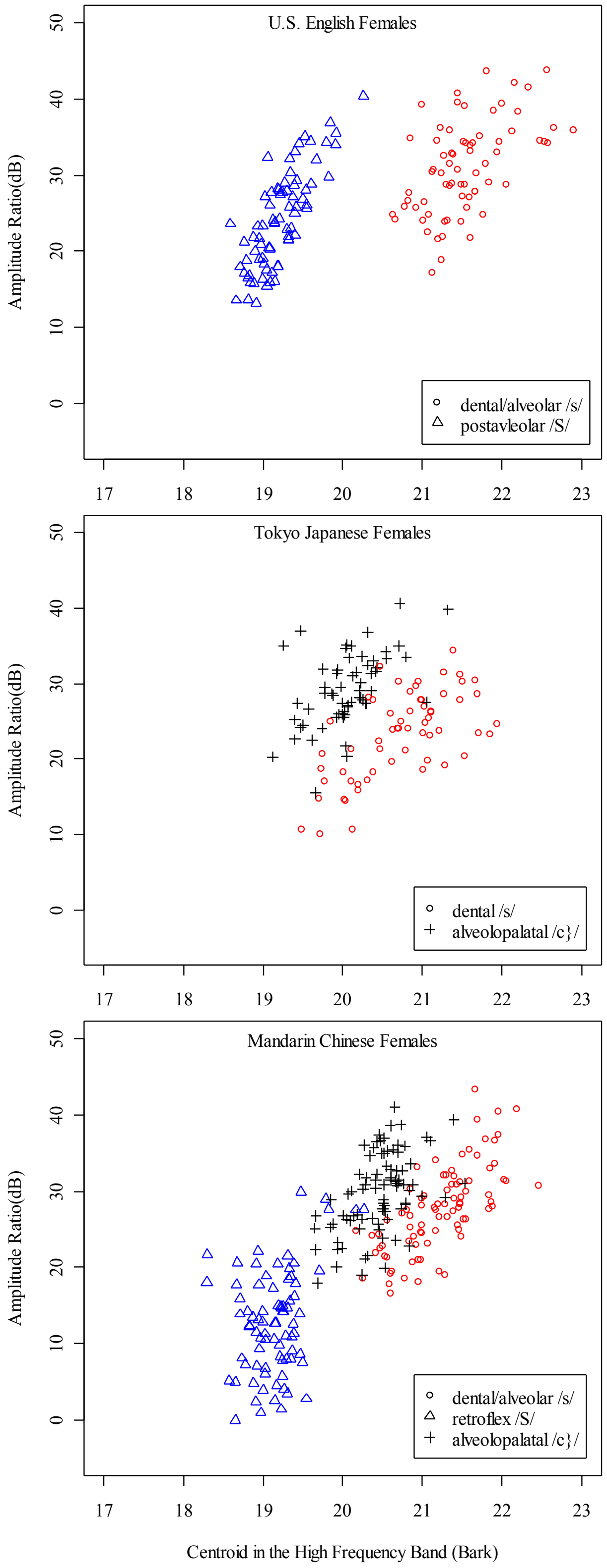
4. DISCUSSION

Our results showed that ampRatio and CentHigh can effectively capture the differences among sibilant fricatives that contrast in two different aspects of articulatory gestures across the three languages. These parameters also showed robust invariance across vowel contexts and speakers, as demonstrated by the fairly clear separation between fricatives in each of the panels in Fig. 2, where the data are shown for all contexts of five vowels and for all five female speakers.

One thing to note is that the /s/ phonemes in these three languages differ in their phonetic details. For example, the English /s/ has a higher centroid than the /s/ in either of the other two languages, whereas the Japanese /s/ has the lowest value of ampRatio among the three /s/ phonemes. A tentative explanation for these differences invokes the notion of maximizing contrast [12]. In English, /s/ contrasts with /ʃ/ primarily in place; and having a higher centroid value makes this place contrast more salient. The especially low value of ampRatio for Japanese /s/ reflects a more dento-laminal articulation, which provides a maximal contrast with /ɕ/ in the amplitude dimension that assesses degree of "palatalization".

In order to evaluate this explanation, we have begun to do perception experiments to see how speakers of these three languages (and of other languages with similar contrasts) respond to differences in ampRatio and CentHigh. The results of these perception experiments may lead us to refine the measures, for example by substituting some psychoacoustic measure of relative loudness such as [6] for the simple amplitude ratio. In the meantime, however, we can at least say that the two spectral measures presented here better capture the two articulatory gestures involved in the contrasting sibilant fricatives of English, Japanese, and Mandarin Chinese. Applying these two measures to sibilant fricative productions in the three languages shows that sounds that are all transcribed as /s/ in IPA can be phonetically very different. It also suggests that these systematic phonetic differences are not random, but rather hinge upon the set of contrast(s) with other sounds in each language's consonant inventory.

Figure 2: Comparison of sibilant fricatives in the three languages, plotting ampRatio against CentHigh.



5. REFERENCES

[1] Behrens, S. J. & Blumstein, S. E. 1988. Acoustic characteristics of English voiceless fricatives: A descriptive analysis. *Journal of Phonetics* 16, 295-298.

[2] Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

[3] Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *J. Acoust. Soc.Am.* 84, 115–124.

[4] Funatsu, S. 1995. Cross language study of perception of dental fricatives in Japanese and Russian. *Proc.13th ICPhS*, Stockholm, Vol 4, pp. 124-127.

[5] Fujisaki, H. & Kunisaki O. 1978. Analysis, recognition, and perception of voiceless fricative consonants in Japanese. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-26, No.1, February 1978.

[6] Glasberg, B., Moore, B. C. J. 2002. A model of loudness applicable to time-varying sounds*, J. Audio Eng. Soc.* 50: 331-342.

[7] Halle, M. & Stevens, K. 1997. The postalveolar fricatives of Polish. In: S. Kiritani, H.Hirose & H. Fujisaki (eds.). *Speech Production and Language: In Honor of Osamu Fujimura*.Berlin. Mouton de Gruyter; 177-193.

[8] Hughes, G. W., & Halle, M. 1956. Spectral properties of fricative consonants, *J. Acoust. Soc.Am.* 28, 303-310

[9] Jongman, A., Wayland, R., & Wong, S. 2000. Acoustic characteristics of English fricatives, *J. Acoust. Soc. Am*, 108 (3), 1252 – 1263.

[10] Ladefoged, P., & Maddieson, I. 1986. Some of the sounds of the worlds' languages (Preliminary version). *UCLA Working Papers in Phonetics*.

[11] Ladefoged, P. & Wu, Z. 1984. Places of Articulation: An Investigation of Pekingese Fricatives and Affricates. *Journal of Phonetics* 12 (3): 267-278.

[12] Liljencrants, L. & Lindblom, B. 1972. Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language,* 48, 839-862.

[13] Maekawa, K., & Kikuchi, H. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In J. van de Weijer, K. Nanjo, and T. Nishihara, eds., *Voicing in Japanese*, Mouton de Gruyter, pp.205-228.

[14] McGowan, R. S., & Nittrouer, S. 1988.Difference in fricative production between children and adults: evidence from an acoustic analysis of /sh/ and /s/. *J. Acoust. Soc.Am.* 83, 229-236.

[15] Shadle, C. H. & Mair, S. J. 1996. Quantifying spectral characteristics of fricatives. *Proc. 4th Int. Conf. Spoken Language Processing*, Philadelphia, U.S.A., 1517 – 1520.

[16] Stevens, K.N. 1998. *Acoustic Phonetics*. Cambridge, Mass.: MIT Press.

[17] Stevens, K. N., Li, Z., Lee, C.-Y., & Keyser, J. 2004. A note on Mandarin Fricatives and Enhancement. In G. Fant, H. Fujisaki, J. Cao., & Y. Xu (Eds.), *From Traditional Phonology to Modern Speech Processing*. Beijing: Foreign Language Teaching and Research Press.

[18] Toda, M., & Honda, K. 2003. An MRI-based cross-linguistic study of sibilant fricatives. *Proc. 6th International Seminar on Speech Production*, Manly (Australia), December 6-10, 2003.