Did you say susi or shushi? Measuring the emergence of robust fricative contrasts in English- and Japanese-acquiring children

Jeffrey J. Holliday, Mary E. Beckman, Chanelle Mays

Department of Linguistics, The Ohio State University, USA
{jeffh,mbeckman}@ling.ohio-state.edu, mays.100@osu.edu

Abstract

While the English fricatives /s/ and /ʃ/ can be well-differentiated by the centroid frequency of the frication noise alone, the Japanese fricatives /s/ and /ɕ/ cannot be. Measures of perceived spectral peak frequency and shape developed for stop bursts were adapted to describe sibilant fricative contrasts in English- and Japanese-speaking adults and children. These measures captured both the cross-language differences and more subtle inter-individual differences related to language-specific marking of gender. They could also be used in deriving a measure of robustness of contrast that captured cross-language differences in fricative development.

Index Terms: sibilant fricatives, acquisition, gender marking, English, Japanese

1. Introduction

Voiceless sibilants are the most commonly attested fricative type across languages [1]. They also typically have a high functional load, occurring in many words within a language and often also bearing socioindexical information about talker gender (e.g.,[2]). At the same time, they are articulatorily complex, require a high level of motor control to articulate, and are generally acquired later than nasals and stops [3]. For example, Japanese-speaking children do not master the /s/-/ɕ/ contrast until 6 years [4]. While the English /s/-/ʃ/ contrast is generally mastered earlier, 7-year-old English-speaking children still produce these sounds in ways that differ subtly from ambient adult models [5]. Moreover, in early stages of acquisition, children acquiring either of these languages can make acoustic distinctions that are imperceptible to adult listeners [6]. This can obfuscate the true trajectory of fricative development in children. It is thus crucial to have acoustic measurements that are able to capture even subtle contrasts in the fricative productions of children as well as adults.

Previous studies have demonstrated that English /s/ and /ʃ/ differ robustly in front cavity size. In /s/, the tongue tip typically contacts the lower incisors to make an alveolar place of articulation, whereas the apical postalveolar posture of the /ʃ/ opens up a sublingual cavity [7], [8], [9]. Also, most speakers use lip protrusion to lengthen the cavity further during /ʃ/ [10].

By contrast, the Japanese fricatives /s/ and /ɕ/ do not differ as robustly in front cavity size. The tongue tip is not raised in the production of /ɕ/ and the lips are spread rather than protruded. In articulatory terms, then, the contrast is characterized more by differences in tongue posture, with /s/ having a laminal or apical constriction, and /ɕ/ having the longer "palatalized" constriction that is characteristic of an alevolopalatal [9]. Acoustically, this corresponds to a higher F2 locus reflecting the shorter back cavity that results.

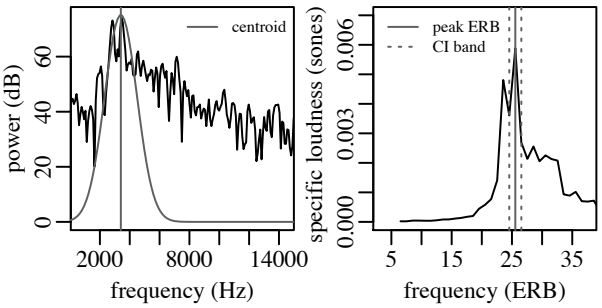

Figure 1: *Acoustic and psychoacoustic measures in a spectrum from the /ʃ/ in English "shoe". See section 2.4 for a more detailed explanation.*

The acoustic difference between English fricatives is often characterized using the spectral mean of the frication noise [11] [12], which is higher in /s/ than in /ʃ/. Unlike vowel formants, however, the fricative centroid is only an indirect measure of the perceptually dominant resonance of the front cavity. It is particularly misleading if there are multiple peaks, as can happen if the constriction is loose and there is back cavity coupling. In these cases the spectral distribution can become bimodal, which pulls the spectral mean downward and hides the true size of the front cavity. Because this hinders our ability to use spectral mean to interpret the articulatory distance between /s/ and /ʃ/, we suggest it may not be an appropriate measure for assessing articulatory development in children.

Another measure commonly used to differentiate fricatives is the frequency of the highest spectral peak [11]. This measure is less sensitive to changes in the shape of the spectrum and is a more robust measure of the actual size of the front cavity. The modification that we suggest is to calculate the peak frequency from an auditory-based ERB-sones spectrum that can reduce some of the noise in higher frequency bands where frequency sensitivity is reduced anyway. That is, because an ERB-sones space is modeled after the human auditory system itself [13], it should do a better job of separating tokens that are meaningfully different (see right panel of Fig. 1).

The measures we are suggesting here were first proposed in [14] to classify stop bursts, although we will be applying them at both the middle and end of the fricative, which will permit a dynamic analysis that is not possible when observing stop bursts. The first measure, peakERB, is the ERB frequency of the loudest spectral peak. The second measure, compactness index (CI), is the proportion of the area of a normalized ERB-sones spectrum that is contained within a 3-ERB band around the peak.

We use these measures in deriving a measure of the robustness of contrast between the two sibilants for individual adult and child speakers in a cross-linguistic cross-sectional study of obstruent development. The measure of robustness of contrast is the rate of accuracy with which a fitted logistic regression model can predict the fricative target. The motivation for this choice of measure is that if tokens of the two fricative categories are robustly differentiated in the psychoacoustic dimensions we have selected, a logistic regression model should be able to accurately predict the fricative target for each token.

2. Methods

The participants in the study were 2- to 5-year-old children and adults who were native speakers of American English recorded in Ohio or native speakers of Japanese recorded in Tokyo (see Table 1).

Table 1: *Number of participants by language and age group.*

English			Japanese		
age	female	male	age	female	male
adult	9	8	adult	10	10
5	12	9	5	8	9
4	10	11	4	12	10
3	10	10	3	13	10
2	9	11	2	10	10

Target fricatives were elicited word-initially (in words such as English *shoe* and Japanese *shumai* 'dumpling') as part of a larger study targeting initial obstruents. Productions were elicited using a picture prompted word-repetition task and transcribed by a native-speaker phonetician [15]. The transcriptions are used here to exclude any misarticulations that were not transcribed as fricative substitutions.

All productions were first marked for fricative onset, indicated by the onset of high-frequency energy in the spectrogram accompanied by a change in amplitude in the waveform. The end of the fricative (i.e. the vowel onset) was marked at the zero crossing on the first upswing after the first clear downswing at the beginning of periodicity in the waveform. The same convention was followed in cases where the signal was fricated even after periodicity began.

We extracted two 8 ms Hamming windows beginning at 90 ms and 10 ms before vowel onset. The two frames for each fricative were then run through a MATLAB script [16] that calculates an ERB-sones spectrum. The peakERB and CI were then calculated for each frame as described in the introduction. For the rest of this paper, measures calculated from the earlier frame will be marked with "f" (i.e. peakERB-f, CI-f), and measures calculated from the frame closer to vowel onset will be marked with "v" (i.e. peakERB-v, CI-v). We expect peakERB-f to correlate inversely with the size of the front cavity and indicate /s/-likeness. We expect peakERB-v to also correlate inversely with front cavity length, but to a lesser degree than peakERB-f because the fricative constriction may already be yielding to the vowel. We expect CI-f to be higher in /s/, reflecting the different effects of the ERB transform on the shape of lower- versus higher-frequency resonance peaks. CI-v, on the other hand, should be lower in /ʃ/, reflecting the lower F2 locus of a longer back cavity. The difference between /s/ and /ɕ/ should be particularly salient due to the convergence of F2 and F3 in the alveolopalatal.

3. Results

3.1. Adult group results

We first built separate models for the adult English and Japanese speakers to see how well peakERB-f and CI-f can differentiate /s/ and /ʃ/ in English and /s/ and /ɕ/ in Japanese. We applied stepwise logistic regressions predicting fricative target with peakERB-f and CI-f as predictor variables to the English- and Japanese-speaking adults' data. We found that the English-speaking adults' targets could be predicted with 79.1% accuracy using only peakERB-f. When separate models were built for each gender the females' targets were predicted with 88.4% accuracy and the males' targets were predicted with 76.6% accuracy, both using only peakERB-f. The Japanese adults' targets were overall predicted with 67.1% accuracy using only peakERB-f, and when separated by gender the females' and males' prediction accuracy rates were 64.9% and 72.3%, respectively. Adding CI-f as a parameter in the group models did not increase prediction accuracy significantly.

The Japanese results suggest that critical information is located elsewhere. We might hypothesize critical information to be in the transition into the vowel, for at least two reasons. First, as noted above, Japanese /s/ and /ɕ/ differ more robustly in back cavity size than in front cavity size. Second, Japanese consonants only occur pre-vocalically and, as [17] points out, the small vowel inventory of Japanese (and the phonotactic restrictions against */si/ and /ɕɛ/) may allow listeners to rely more on vocalic cues than would be feasible in languages with more vowels. In English, on the other hand, the fricative-internal cues must be robust enough to identify the segments in initial, medial, and coda position, as well as in the context of any of the 15 vowels of the languages or any of the consonants that can occur in clusters with the the two sibilant fricatives.

We tested this hypothesis by adding to the stepwise model peakERB-v and CI-v. When the Japanese adult data was run through a stepwise logistic regression using all four parameters the model kept all four, but we chose to remove peakERB-v and CI-f because doing so did not reduce prediction accuracy significantly and we wanted to avoid an overfit. The two-parameter model (peakERB-f and CI-v) predicted the Japanese adults' targets with 74.3% accuracy, and when the genders were run separately the females' and males' targets were predicted with 72.2% and 78.2% accuracy, respectively. These results suggest that the "community norm" model for English involves a simple contrast localized in the fricative itself, whereas the community norm for Japanese involves a dynamic complex of cues that are spread over the CV mora as a whole.

3.2. Individual adult results

Next we built separate stepwise logistic regression models for each speaker to explore patterns of inter-individual variation within each language. That is, two models were built for each speaker's data: a best-fit community norm model and a best-fit individual model. In cases where the stepwise regression chose a model with more than two parameters, we removed parameters until we got the best fit with only two parameters.

Accuracy rates when the community norm model was applied to the English speakers were quite high overall, with a mean accuracy of 89.5%, and five out of the 17 showing perfect discrimination. For eight subjects, the best individual model was the community norm model, and for all of the others, the best individual model either used CI-f alone (two subjects) or used both fricative-internal measures (seven subjects) (see Ta-

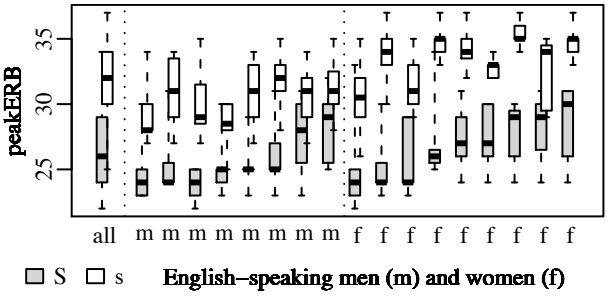

Figure 2: *peakERB-f values of individual adult English speakers. /ʃ/ (="S") values do not vary as much between genders, but females appear to have higher values for /s/, which would contribute to the contrast being more robust.*

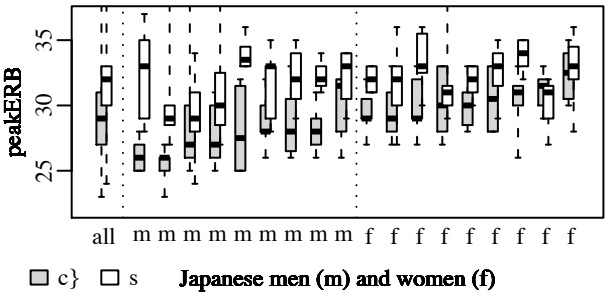

Figure 3: *peakERB-f values of individual adult Japanese speakers. Males tend to have lower values for /ɕ/ (="c}") than females do, but values for /s/ appear relatively stable across genders.*

ble 2). Females showed a high prediction accuracy on average, at 94.4%, with males at 84.1%. Figure 2 suggests an explanation for the more robust differentiation in women. Six of the eight women had very high median peakERB-f in /s/. This is reminiscent of findings by [2] and [18], suggesting that a fronter /s/ is a gender marker for women.

The models for the Japanese-speaking adults reveal three trends. First, there was much less homogeneity among the individual speakers (see Table 2). Second, vocalic transition information (i.e. a "v"-measure) can be helpful in defining the /s/-/ɕ/ contrast. Nine of the Japanese adult speakers chose one of these parameters for their model. These results are consistent with our hypothesis that fricative-internal information is less robustly contrastive in Japanese, and also agree with the Japanese adult group results reported above. In keeping with this, Figures 3 and 4 show much less separation between Japanese /ɕ/ and /s/ than Figure 2 shows for /ʃ/ and /s/. Third, Japanese women do not show a more robust differentiation between the fricatives than Japanese men. If anything, they show less robust separation with the individual models yielding mean accuracy for females at only 78.8%, as compared to males at 89.8%.

3.3. Individual children's results

The individual models for the children were built in the same way as the individual adult models. The English-acquiring children's models were allowed to choose from peakERB-f and CI-f, and the Japanese-acquiring children's models were allowed to choose from peakERB-f, peakERB-v, CI-f, and CI-v.

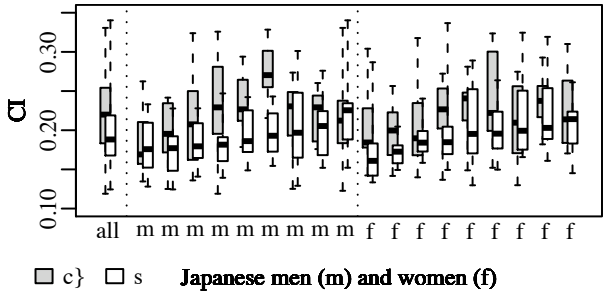

Figure 4: *CI-v values of individual adult Japanese speakers. There does not appear to be any gender difference, but the CI-v does help differentiate the two sibilants for most speakers.*

The English-speaking children were like the English-speaking adults in that they showed a high degree of homogeneity. Of the 65 children who had any contrast, 38 used peakERB-f alone, and only 7 used CI-f alone. What does differentiate the children from the adults is how well the model can predict the speakers' fricative targets. For 13 of the children, in fact, the best fit was an intercept model, indicating that none of these cues differentiates the fricatives. Moreover, looking at Table 3 we see a large drop in prediction accuracy from ages 5 to 4 and 3 to 2.

A second salient result was a large difference in robustness of contrast between boys and girls. From age 3 on, the trend line in Figure 5 for girls shows a much more robust contrast than for boys. A plausible explanation for this difference could be the acquisition of adult English speakers' patterns of gender marking discussed above.

Examination of the Japanese-acquiring children supports this explanation. There is no difference between male and female 5- or 4-year-old children in the robustness of contrast. As a consequence, Japanese-acquiring children were overall less accurate than English-acquiring children. Japanese-acquiring children also differed from English-acquiring children in showing much less homogeneity in which cues were used. This lack of homogeneity is in keeping with the adult patterns.

Table 2: *Parameters chosen in individual models.*

	English		Japanese	
types	Adults	Children	Adults	Children
pf	8	38	2	4
pf+cf	7	20	5	14
cf	2	7	1	4
pf+pv	0	0	2	2
pf+cv	0	0	3	5
cf+pv	0	0	0	6
cf+cv	0	0	2	3
pv	0	0	1	4
cv	0	0	1	1
pv+cv	0	0	0	3
1	0	13	1	11

What is consistent across both languages is that children's fricative contrasts become more robust as the children develop. In Figure 5 it is clear that the measure of robustness is positievly correlated with age in both language groups. Although the Japanese-acquiring children seem to have overall less ro-

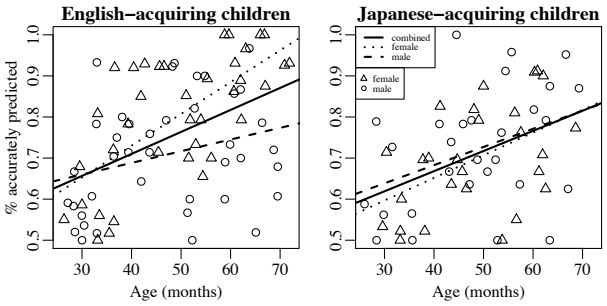

Figure 5: *Increase in robustness of contrast in the individually fit models against age in months. Note the increase in gender differences in English.*

bust contrasts than the English-acquiring children, it is interesting to note in Table 3 that Japanese 5-year-old females have a robustness of contrast measure nearly identical to Japanese adult females. Thus it is important to bear in mind that adult-like performance, and not necessarily 100% differentiation, is the end-state of acquisition.

Table 3: *Summary of robustness of contrast based on individually fit models across languages and age groups.*

	English			Japanese		
Age	All	Female	Male	All	Female	Male
adult	0.931	0.974	0.883	0.843	0.788	0.898
5	0.835	0.910	0.745	0.780	0.784	0.776
4	0.775	0.812	0.742	0.729	0.730	0.728
3	0.783	0.823	0.740	0.736	0.708	0.771
2	0.611	0.600	0.618	0.592	0.574	0.604

4. Conclusions

In this paper we adapted psychoacoustic measures developed for differentiating lingual stop bursts to the description of sibilant fricative contrasts in English and Japanese. We calculated peakERB and CI both in the middle of the fricative and more toward the transition into the vowel. These measures captured known cross-language differences between the sibilant fricatives of English and Japanese. They were also sensitive enough to capture inter-individual variation in the degree to which different aspects of these sounds differentiated the two types. Examination of fricative-internal aspects revealed evidence for gender marking of /s/ by some Ohio English-speaking women. The measures were also amenable for use in deriving a measure of degree of robustness of contrast in individual children. Correlating the robustness of contrast measure with age revealed cross-language differences in the trajectory of fricative development in English- and Japanese-acquiring children. While the youngest children showed comparably undifferentiated fricatives in both languages, the increase in robustness with age was steeper in English-acquiring children. It was particularly steep in English-acquiring girls, who seem to be learning the gender marking patterns of the ambient speech community.

5. Acknowledgements

Data collection and analysis was supported by grants NIDCD 02932 to Jan Edwards and NSF BCS-0729306 to Mary Beckman. Thanks to Tim Arbisi-Kelm and Eun Jong Kong for their work on developing the psychoacoustic measures and to Julie Johnson, Fangfang Li, Oxana Skorniakova, and Asimina Syrika for help in data preparation.

6. References

[1] Ladefoged, P., Maddieson, I., "The Sounds of the World's Languages", Blackwell, MA (1996).

[2] Heffernan, K., "Evidence from HNR that /s/ is a social marker of gender", Toronto Working Papers in Linguistics 23.2: 71-84 (2004).

[3] Kent, R., "The biology of phonological development", In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), Phonological development: Models, research, implications (pp. 65.90). Timonium, MD: York Press (1992).

[4] Nakanishi, Y., Owada, K., Fujita, N., "Koon kensa to sono kekka no kosatsu [Results and interpretation of articulation tests for children]". RIEEC Report [Annual Report of Research Inst. Education of Exceptional Children, Tokyo Gakugei Univ.], 1, 1.41 (1972).

[5] McGowan, R., Nittrouer, S., "Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/", *JASA* 83:1, 229-236 (1988).

[6] Li, F., Edwards, J., Beckman, M. E., "Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers", *J.Phon.* 37, 111-124 (2009).

[7] Perkell, J., Boyce, S., and Stevens, K., "Articulatory and acoustic correlates of the [s-š] distinction", In Wolf, J. J. and Klatt, D. H. (eds.) Speech Communication Papers presented at the 97th meeting of the Acoustical Society of America. American Institute of Physics, New York: 109-113, (1979).

[8] Narayanan, S., Alwan, A., Haker, K.,"An articulatory study of fricative consonants using magnetic resonance imaging", *JASA* 98:3, 1325-1347 (1995).

[9] Toda, M., Honda, K., "An MRI-based cross-linguistic study of sibilant fricatives", Proc. of the 6th Int'l Seminar on Speech Prod. (2003).

[10] Toda, M., Maeda, S., Carlen, A. J., Meftahi, L., "Lip gestures in English sibilants: articulatory-acoustic relationship, proc. 7th ICSLP, 2165-2168, (2002).

[11] Jongman, A., Wayland, R., Wong, S., "Acoustic characteristics of English fricatives", *JASA* 108:3, 1252-1263 (2000).

[12] Nissen, S. L., Fox, R. A., "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective", *JASA* 118, 2570-2578 (2005).

[13] Glasberg, B. R., Moore, B. C. J., "Derivation of auditory filter shapes from notched-noise data", *Hearing Research* 47:1-2, 103-138 (1990).

[14] Arbisi-Kelm, T., Beckman, M. E., Kong, E-J., Edwards, J., "Psychoacoustic measures of stop production in Cantonese, Greek, English, Japanese, and Korean", Poster presented at the 156th Meeting of the Acoustical Society of America, Miami, 10-14 November (2008).

[15] Edwards, J., Beckman, M. E., "Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes", Language Learning & Development, 4(2): 122-156 (2008).

[16] Timoney, J., Lysaght, T., Schoenwiesner, M., McManus, L, "Implementing loudness models in MATLAB", Proc. of the 7th Int. Conference on Digital Audio Effects, Naples, Italy (2004).

[17] Toda, M., "Speaker normalization of fricative noise: Considerations on language-specific contrast", Proceedings of ICPhS XVI, 825-828 (2007).

[18] Stuart-Smith, J., Timmins, C., Wrench, A., "Sex and gender differences in Glaswegian /s/", Proceedings of ICPhS XV, 1851-1854 (2003).