Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development

Fangfang Li[a)]
Department of Psychology, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1J 3M4, Canada

Benjamin Munson
Department of Speech-Language-Hearing Sciences, University of Minnesota, 164 Pillsbury Avenue South East, Minneapolis, Minnesota 55455-0000

Jan Edwards
Department of Communicative Disorders, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705

Kiyoko Yoneyama
Department of English, Daito Bunka University, 1-9-1 Takashimadaira, Itabashi, Tokyo, Japan 175-8571

Kathleen Hall
Department of English, College of Staten Island, City University of New York, 2S–218, 2800 Victory Boulevard Staten Island, New York 10314

(Received 9 April 2010; revised 18 October 2010; accepted 22 October 2010)

Both English and Japanese have two voiceless sibilant fricatives, an anterior fricative /s/ contrasting with a more posterior fricative /ʃ/. When children acquire sibilant fricatives, English children typically substitute [s] for /ʃ/, whereas Japanese children typically substitute [ʃ] for /s/. This study examined English- and Japanese-speaking adults' perception of children's productions of voiceless sibilant fricatives to investigate whether the apparent asymmetry in the acquisition of voiceless sibilant fricatives reported previously in the two languages was due in part to how adults perceive children's speech. The results of this study show that adult speakers of English and Japanese weighed acoustic parameters differently when identifying fricatives produced by children and that these differences explain, in part, the apparent cross-language asymmetry in fricative acquisition. This study shows that generalizations about universal and language-specific patterns in speech-sound development cannot be determined without considering all sources of variation including speech perception. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3518716]

PACS number(s): 43.71.Hw, 43.70.Ep, 43.70.Kv, 43.71.Gv [AJ] Pages: 999–1011

I. INTRODUCTION

A. Overview

It has long been recognized that children's first words deviate somewhat from those produced by the adults to whom they are exposed during acquisition. Children's early productions frequently demonstrate omission and substitution errors relative to the adult forms. Many of these errors appear to be fairly consistent across children and across languages. For example, it has been observed across many languages that children produce vowels earlier than consonants and that they produce certain consonants, such as stops, earlier than others, such as fricatives or affricates. Jakobson (1941/1960) termed these cross-linguistically invariant sound acquisition sequences "implicational universals" and suggested that these regularities reflect principles that drive the organization of adult sound systems of human languages as well as children's speech development. In this view, the earlier acquisition of stop consonants relative to other consonants would be the evidence that stops are universally "easier" to acquire than other consonants. Jakobson further pointed out that within stops, the sounds produced further back in the oral cavity, such as /k/, usually occur later and are replaced by the production of more front ones, such as /t/, and Locke (1983) termed this as the *fronting* universal and extended it to the class of fricatives, arguing that the anterior sibilant fricative /s/ is universally easier than its post-alveolar counterpart, /ʃ/.

The hypothesis that fronting is a universal pattern in child language acquisition is not supported by cross-language studies of fricative acquisition. One notable example is the difference in error patterns in the acquisition of voiceless sibilant fricatives in English and Japanese (Beckman *et al.*, 2003; Li *et al.*, 2009). Both languages contrast an anterior voiceless sibilant fricative /s/ with a more posterior fricative /ʃ/. Large-scale normative studies report more fronting errors, i.e., [s]-for-/ʃ/ substitutions, in English-acquiring children, but more backing errors, i.e., [ʃ]-for-/s/ substitutions, in Japanese-acquiring

[a)]Author to whom correspondence should be addressed. Electronic mail: fangfang.li@uleth.ca

children. Specifically, Sander (1972) used data from normative studies of the acquisition of English by Wellman *et al*. (1931) and Templin (1957) and determined that the average age of acquisition for /s/ is 3 yr, 0 months and for /ʃ/ is 4 yr, 0 months, using the criterion of correct use of the speech sound in more than two word positions in over 50% of the children being tested. Similarly, Smit *et al*. (1990) examined speech-sound acquisition in 117 English-speaking children aged 3 to 9 yr and also found that /s/ is acquired at the age of 3 yr, 0 months in word-initial position, whereas word-initial /ʃ/ is acquired at 4 yr, 0 months. In contrast, Yasuda (1970) studied 100 Japanese-speaking children aged 3 yr, 0 months to 3 yr, 11 months and found that production accuracy for /ʃ/ (60.3%) is much higher than that for /s/ (24.5%). These consonants were investigated only in word-initial and word-medial positions, as Japanese has a restricted distribution of word-final consonants.

It is important to note that the primary method used in these large normative studies was phonetic transcription by native speakers. This presumes that children articulate speech sounds in a manner similar to adults and their productions can therefore be accurately placed into adults' perceptual categories. This assumption has been seriously challenged by the instrumental analysis of children's speech. Mounting evidence has shown the existence of distinctive sound productions by children that are well within the perceptual boundary of a single sound category of adults, a phenomenon termed "covert contrast" (see Scobbie *et al*., 2000, for a review). For example, in an electropalatography (EPG) study, Gibbon *et al*. (1995) have found more retracted lingual-palatal contact for /ʃ/ than /s/ targets, even when transcribers described them as homophonous lateral fricatives [ɬ].

Another limitation of the transcription method lies in a possible constraint from transcribers' language-specific knowledge. It has been well established that language-specific perceptual knowledge biases listeners' perception of unfamiliar foreign-language speech sounds (Best, 1990, 1995; Best and Tyler, 2007; Iverson and Kuhl, 1995; Pierre and Best, 2007). These biases emerge when children's speech perception becomes tuned to the language they are acquiring, typically around the end of the first year of life (Best and McRoberts, 2003; Best *et al*., 1988; Kuhl *et al*., 1992; Nittrouer and Lowenstein, 2010; Werker and Lalonde, 1988; Werker *et al*., 1998). However, little attention has been paid to how adult listeners' perception of children's speech is constrained by language-specific phonological knowledge. As Scobbie (1998) points out: "We should not forget that from the perspective of adult ears, the speech of all infants is another example of the 'unfamiliar'" (p. 343). The traditional transcription method relies on auditory impressionistic judgments and is likely to introduce perceptual biases to the description of children's early immature speech. One example of this is given in Edwards and Beckman's (2008) study of cross-linguistic differences in speech-sound acquisition. They observed that two Greek-speaking trained phonetic transcribers denoted some young Greek-speaking children's productions of target /ki/ as correct, while similarly trained English-speaking phonetic transcribers labeled the same productions as [ti]-for-/ki/ substitutions. This suggests the existence of fine-grained cross-linguistic differences in perception. Consequently, it is not easy to determine whether language-specific acquisition patterns, such as fricative acquisition in English and Japanese, are due to cross-linguistic differences in children's speech production or due to cross-linguistic differences in how adults perceive children's speech. The asymmetries in fricative development in these two languages may provide counter evidence to the hypothesis that there is a universal order of acquisition for fricatives, or it may be evidence of an adult perception bias introduced during transcription, which obscures a universal pattern. The current study is an effort to evaluate the possible effect of the latter, that is, how language-specific perception affects the identification of errors in children's speech.

B. Language-specific articulation and acoustics of voiceless sibilant fricatives

One reason to suspect that English and Japanese speakers would perceive children's fricatives differently is the subtle difference between these shared sounds in the two languages, both with respect to their articulation and to their acoustics. First consider the anterior fricatives, which are transcribed as /s/ in both languages. The English /s/ is an apico-alveolar sound, whereas the Japanese /s/ is more of a laminal-dental sound (Akamatsu, 1997). Moreover, the Japanese /s/ has also been shown to be less intense and less sibilant than the English /s/, which presumably reflects a more distributed spectrum in the acoustics (Akamatsu, 1997). The posterior sibilant fricatives in the two languages differ even more, such that there is some controversy as to whether the two posterior fricative sounds in English and Japanese should be denoted with the same phonetic symbol at all. In many early studies, the Japanese post-alveolar sibilant fricative was transcribed as /ʃ/ (Funatsu, 1995; Nakata, 1960). More recent studies, such as Ladefoged and Maddieson (1996) and Toda and Honda (2003), suggest that the Japanese post-alveolar sibilant has a distinct enough articulatory configuration from English /ʃ/ to warrant using a different symbol, /ɕ/. Particularly, English /ʃ/ is produced with the tongue blade retracted and raised to form a narrow constriction in the oral cavity (Narayanan *et al*., 1995), whereas the Japanese post-alveolar fricative is produced with the tongue's pre-dorsum region bunched up to form a palatal channel above the tongue (Toda and Honda, 2003). Furthermore, English /ʃ/ is produced with rounded lips (presumably to increase the size of the resonant cavity anterior to the constriction, thereby increasing the concentration of energy in the lower frequencies and enhancing the contrast between /ʃ/ and /s/), but the Japanese post-alveolar is not. Nonetheless, the two sounds are sufficiently comparable across the two languages that they can be readily assimilated into the other language (e.g., narrowly transcribed Japanese [sɯɕi] is perceived as [suʃi] in English; English [ʃɑk] is perceived as [ɕokkɯ] in Japanese). Furthermore, because the primary phenomenon of interest in this study is children's substitution errors, and the symbols /ʃ/ and /s/ are sufficient to show the direction of the substitution error (i.e., whether the error

is fronting or backing) equally well for both languages, we will use the /ʃ/ symbol for both English and Japanese.

A wealth of studies has examined how the English voiceless sibilant fricatives are differentiated from one another acoustically. Most of these studies suggest that the two voiceless sibilants can be differentiated by the spectral properties of the frication alone (Behrens and Blumstein, 1988; Hughes and Halle, 1956; Jongman *et al*., 2000). This is because English /s/ and /ʃ/ differ primarily in the major lingual constriction in the oral cavity, with the place of the constriction being further back in /ʃ/ than in /s/. The fricative noise spectrum principally reflects resonances in front of the major constriction that are further enhanced by rapid air stream impinging on the incisors (Fant, 1960; Shadle, 1991; Stevens, 1998). Hence, retracting the tongue further back in producing /ʃ/ results in a longer front cavity, which then lowers the overall frequency range in the major energy concentration of the noise spectrum.

These differences between English /s/ and /ʃ/ can be captured by a widely used technique for describing spectral properties of fricatives, spectral moments analysis. This analysis treats the fricative noise spectrum as a probability density distribution and calculates the statistical moments of the distribution (Forrest *et al*., 1988). The first moment (henceforth, M1), also called *centroid frequency*, is the mean frequency of the spectral energy distribution in the noise spectrum and is negatively correlated with the length of the front cavity. The longer the front resonating cavity is, the lower the overall resonating frequencies in the fricative spectrum will be, which is reflected in a lower M1 value. Therefore, the M1 value of /s/ is expected to be higher than that of /ʃ/ because of the shorter front resonating cavity in /s/. This prediction has been confirmed robustly in many acoustic studies of English fricatives (Forrest *et al*., 1988; Jongman, *et al*., 2000; Nissen and Fox, 2005; Nittrouer, 1995; Shadle and Mair, 1996; Fox and Nissen, 2005).

There are three other moments that spectral moments analysis computes: *standard deviation* (the second moment, henceforth M2), *skewness* (the third moment, henceforth M3), and *kurtosis* (the fourth moment, henceforth M4), each of which describes a different dimension of the fricative spectral shape. Specifically, M2 calculates how much the spectrum energy deviates from the centroid frequency and thus provides an index of variance; M3 computes the energy difference above and below the centroid frequency in order to capture the overall shape of the spectral distribution; and M4 measures the peakedness of the fricative energy distribution relative to the normal distribution. Jongman *et al*. (2000) examined English fricatives in 20 English-speaking adults using these four spectral moments and found that M1, M3, and M4 are able to distinguish /s/ from /ʃ/. In a more recent study, Li *et al*. (2009) examined English voiceless sibilant fricatives using a mixed effects model including all four spectral moments as predictors and found that M1 is the primary acoustic correlate for the /s/-/ʃ/ contrast and M1 by itself is sufficient to distinguish the two fricatives once individual differences have been accounted for. Nittrouer (1995) also applied moments analysis to fricative productions by English-speaking children aged 3, 5, and 7 yr as well as by the adults. She found age-related differences in M1 and M3. Specifically, the difference in M1 between children's /s/ and /ʃ/ is smaller than that of adults, suggesting less precise articulatory gesturing in children's production of these voiceless sibilant fricatives. Miccio *et al*. (1996) found all four moments are effective in describing the /s/-/ʃ/ distinctions produced by normal developing children. Similarly, Nissen and Fox (2005) and Fox and Nissen (2005) also utilized spectral moments analysis to describe fricative productions by children, adolescents, and adults. They found that all four moments are useful in describing children's /s/ and /ʃ/ distinctions and the two sounds are better distinguished acoustically as children's ages increase.

Relatively few studies have described the acoustic characteristics of Japanese voiceless sibilant fricatives. Funatsu (1995) examined the acoustics of the Japanese /s/-/ʃ/ contrast and the Russian /s/-/sʲ/-/ʃ/ contrast. He found that the main peak frequency in the fricative noise (i.e., the frequency that is the most intense) along with the frequency of the second formant of the following vowel at its onset (henceforth *onset F2 frequency*) are sufficient to describe the fricative contrasts in both the languages. Onset F2 frequency has been shown to correlate negatively with the length of the back resonating cavity (Halle and Stevens, 1997; Stevens *et al*., 2004). Because the production of Japanese /ʃ/ involves a dome-shaped tongue posture that creates a long palatal channel, which effectively shortens the length of the back cavity, the value of onset F2 frequency is higher for /ʃ/ than for that for /s/. Li *et al*. (2009) compared the acoustic differences in the voiceless sibilant fricative contrast in Japanese-speaking adults and children and found that M1, onset F2 frequency, and M2 are needed to differentiate the two fricatives in Japanese. The differences in articulation between the two pairs of voiceless sibilants in the two languages, as well as evidence from acoustic studies, lead us to predict that English and Japanese speakers will be likely to use different acoustic cues in identifying voiceless sibilant fricatives, including those produced by children.

C. Language-specific perception of voiceless sibilant fricatives

Much of the research on the perception of fricatives has focused on the relative contribution of information in the frication and the vowel to listeners' identification. Harris (1958) cross-spliced fricative noise portions of /s/ and /ʃ/ with the vocalic portions taken from /s/- and /ʃ/-initial words and found that English-speaking listeners' labeling is more strongly influenced by fricative-internal information (i.e., M1) than information in formant transitions (i.e., onset F2 frequency). Similar results were obtained by LaRiviere (1975). Subsequent studies such as Whalen (1984, 1991) using synthetic speech have shown that fricative-vowel transitions also play an important role in differentiating the /s/-/ʃ/ contrast in English. Moreover, Nittrouer (1992) found that the weight that listeners assign to fricative noise characteristics over fricative-vowel transitions changes as a function of age. In a series of studies, Nittrouer and coworkers combined both synthetic and natural fricative noise with F2 transitions

from different vowels and found that adults differ from children in that they rely more heavily on fricative-internal cues for the /s/-/ʃ/ contrast, whereas children assign more weight to the transitional cue in their perception (Nittrouer, 1996, 2002; Nittrouer and Miller, 1997).

Fewer studies have examined Japanese speakers' perception of Japanese voiceless sibilant fricatives. Nakata (1960) evaluated Japanese listeners' judgments of synthetic fricatives and found that the change of the percept from /s/ to /ʃ/ is primarily correlated with the decrease in resonant frequency of the fricative noise spectrum. He also found that the F2 locus and the relative intensity of the fricative and the following vowel are important in accounting for Japanese listeners' fricative judgments, although the effects of these two cues are not as pronounced as the fricative-internal cue. Another study was conducted by Hirai *et al.* (2005) who examined 42 native Japanese adults' fricative perception using a procedure similar to that used by Nittrouer and colleagues. Hirai *et al.* found that most Japanese adults give more weight to the fricative noise spectrum cue than to the formant transition cue, in a manner similar to the English-speaking adults tested in the work by Nittrouer. However, a small number of adults showed a different weighting strategy in which transitions override the fricative noise information.

The studies cited thus far have all used adult speech as stimuli or synthetic stimuli modeled on the characteristics of adult speech. The variability in these stimuli is either limited (in studies using natural-speech produced by adults) or planned and carefully controlled (in studies using synthetic speech). Aoyama *et al.* (2008) conducted a study that examined the perception, by 12 English-speaking judges, of natural productions of L2 (English) words beginning with /s/ and /θ/, produced by both Japanese-speaking adults and children. They found that target /s/ productions were identified as such with an accuracy of 89% or greater, with most errors labeling productions as /θ/.

D. Purposes

The current paper reports on an experimental paradigm similar to that used by Aoyama *et al.* (2008) but with a focus on a different contrast (specifically, the /s/-/ʃ/ contrast) to examine cross-linguistic differences in adults' perception of children's speech. Particularly, we test adults' perceptions of voiceless sibilant fricatives using children's speech, in order to assess whether cross-linguistic differences in adults' perception of the voiceless sibilant fricative contrast might explain—at least in part—the previously reported cross-language asymmetries in the acquisition of these sounds in English and Japanese. Moreover, by using natural productions from adults and children, our listeners were presented with the natural sources of variability that are present in actual speakers' productions. This allows us to examine statistically the extent to which adults are affected by all of the variation that is present in natural productions, including not only variation in the parameters known to best differentiate between target productions (here, M1 and onset F2 frequency) but also all of the other parameters we measured (M2, M3, and M4). In a sense, our use of this variation gives us a natural-speech analog to the synthetic speech continua used in many perception experiments: The adult speech tokens serve as the best exemplars of a category (i.e., the endpoints), and the children's speech forms a natural, multidimensional continuum between those clear endpoints.

Based on the articulatory and acoustic differences in adult productions of voiceless sibilant fricatives in the two languages, we predicted that adult native listeners of English and Japanese would parse the multidimensional acoustic space differently, especially for children's productions that were not clear exemplars of these sounds. A finding that Japanese-speaking listeners are biased to perceiving productions as /ʃ/ and that English-speaking listeners are biased to perceiving these same productions as /s/ would suggest that the apparent cross-linguistic asymmetry in acquisition of these sounds is attributable in part to cross-linguistic differences in adults' perception of children's speech.

II. METHODS

A. Stimuli

1. Stimulus selection

The stimuli were consonant-vowel sequences excised from real words produced by 2- to 3-year-old children acquiring English or Japanese as a first language. They were elicited using a picture-prompted auditory word-repetition paradigm and were collected as part of a larger project that examined children's phonological development across different languages (Edwards and Beckman, 2008). The stimuli were taken from productions of words with target /s/ and target /ʃ/. The number of syllables each word contains was varied in order to elicit words that are familiar to children. The majority of English words are monosyllabic, and the majority of Japanese words are disyllabic with the primary stress on the first syllable. The target phoneme always occurs in word-initial position. For a complete list of words from which the stimuli were selected, please refer to Li *et al.* (2009). Also, Edwards and Beckman (2008) discussed in detail the effect of all of the stimulus characteristics including word length, prosodic pattern, etc. Productions of 41children were included in the stimuli. Table I lists the breakdown of the speakers in terms of language and age. Stimuli included productions transcribed as being correct and ones transcribed as containing the substituted fricatives with either [s] for /ʃ/ or [ʃ] for /s/. Words whose initial fricatives were transcribed as having stopping errors or other fricative substitution errors (i.e., [f] or [θ] substitutions) were excluded. For each language, all stimulus items were transcribed first by an experienced native-speaker phonetician. A second native-speaker phonetician independently transcribed

TABLE I. Number of participants contributing to the stimuli used in the perception experiments.

	English	Japanese
2-year-olds	9	10
3-year-olds	13	8
Adults	3	3

20% of the data. Phoneme-by-phoneme inter-rater reliability was 90% for English-speaking children and 89% for Japanese-speaking children. Furthermore, as shown in Table I, the stimulus set also contained some productions from adults who were recorded in a word-repetition task and whose recordings were made as potential audio prompts for the repetition task used to elicit children's productions. The purpose of including adult tokens was to ensure that listeners also heard clear adult exemplars of the target sounds, in addition to the children's productions.

A total of 400 consonant-vowel (CV) stimuli were selected. Two hundred tokens from English-speaking children and adults and 200 tokens from Japanese-speaking children and adults were used. Within each language, children's productions were selected based on the native-speaker transcriptions. Specifically, for English-speaking children, 50 tokens of correct /s/ productions, 50 tokens of correct /ʃ/ productions, and 50 tokens of [s]-for-/ʃ/ substitutions were selected. Because the error patterns are extremely skewed so that there were only a few [ʃ]-for-/s/ substitutions in the English database, eight tokens of [ʃ]-for-/s/ substitutions were selected to reflect the true skewed error patterns between the two targets in the database. The remaining 42 English tokens were adult productions. The 200 Japanese tokens were selected based on similar principles, except that there were 50 tokens of [ʃ]-for-/s/ substitutions and only 11 [s]-for-/ʃ/ substitutions because of the opposite error patterns for English- and Japanese-speaking children. In addition, within each transcription category, vowel context and the gender and age of the speakers were balanced as much as possible. All stimuli were normalized for amplitude, and cosine-squared off-ramping was used to minimize acoustic artifacts resulting from extraction.

Five spectral parameters were applied to measure the acoustic characteristics of the speech stimuli. These spectral measures included the first four moments of a spectral moments analysis, which describe the fricative-internal characteristics (hereafter, M1–M4), and the onset F2 frequency of the vowel immediately following the fricative. Li *et al.* (2009) provide a comprehensive description of how these acoustic parameters were obtained. PRAAT (Boersma and Weenink, 2005) was used to segment frication noise and to extract various acoustic parameters. The beginning of frication was defined as the first appearance of aperiodic noise evident both in the sound waves and in the spectrograms. The onset of the vowel that follows the target fricative was identified as the first periodic pulse in the wave form, where onset F2 was measured. The values of the four moments were calculated on fast Fourier transform (FFT) spectra over a 40-ms window that was centered in the frication noise. The distribution of the English and the Japanese stimuli in the five acoustic dimensions is shown quantitatively in Fig. 1. In the dimensions of M1, M3, and M4 and onset F2, the stimuli of both languages show Gaussian-like distributions, with the values for /s/ and /ʃ/ overlapping with each other. For M2, Japanese stimuli have a higher mean value than the English stimuli. Moreover, the English stimuli exhibit a bimodal distribution in the M2 dimension with some stimuli having a higher M2, with a mean around 1500 Hz than others, which have a mean around 500 Hz. A closer examination of the

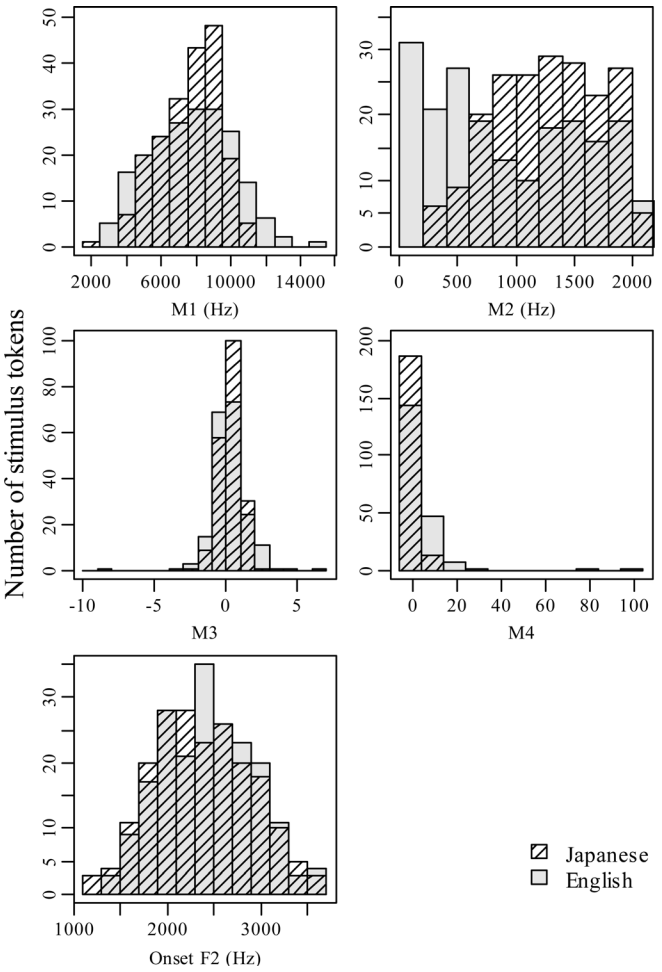

FIG. 1. Distributions of English vs Japanese stimuli on the five acoustic dimensions including the four spectral moments and onset F2 frequency.

nature of this bimodality reveals that the clustered stimuli with lower M2 mode are mostly adults' productions, whereas those of higher M2 mode are all children's productions.

2. Participants and task

Nineteen English-speaking adults were tested in Minneapolis, MN, and 20 Japanese-speaking listeners were tested in Tokyo, Japan. All participants had normal speech, language, and hearing based on self-report. None of the speakers were bilingual, although all of the Japanese speakers had studied English in school and all of the English speakers had studied a second language as part of their university requirements.

The task was speeded classification. Each listener heard 2 blocks of the same 400 tokens. The English and Japanese CV sequences were combined in a single block, and listeners were not told that they were listening to productions from two languages. In one block, listeners were asked "Is it an 's'?" and in the other block, listeners were asked "Is it an 'sh'?" Orthography appropriate to the two languages was used. For example, in English, the word-initial consonant was described either as "'s,' the first sound in *see*, *say*, *sock*, *sew*, *Sue*" or as "'sh,' the first sound in *she*, *shape*, *shock*, *show*, *shoe*." It should be noted that the instructions for

TABLE II. Summary of <s> and <sh> perceptions (as gauged by agreement by more than 70% of all the English-speaking listeners or all the Japanese-speaking listeners) as a ratio to the intended /s/ or /ʃ/ target by the two listener groups. The raw counts of stimulus for each category are included in parentheses.

		English listeners ($n = 18$)		Japanese listeners ($n = 20$)	
		<s>	<sh>	<s>	<sh>
English stimuli (%)	Intended /s/ ($n = 78$)	65 ($n = 51$)	7 ($n = 6$)	43 ($n = 34$)	6 ($n = 5$)
	Intended /ʃ/ ($n = 122$)	20 ($n = 25$)	48 ($n = 59$)	12 ($n = 15$)	43 ($n = 53$)
Japanese stimuli (%)	Intended /s/ ($n = 118$)	40 ($n = 47$)	12 ($n = 14$)	40 ($n = 47$)	13 ($n = 15$)
	Intended /ʃ/ ($n = 82$)	11 ($n = 9$)	55 ($n = 45$)	12 ($n = 10$)	30 ($n = 25$)

English-listeners are straightforward as the labels "s" and "sh" are transparent from the orthography. For Japanese listeners, the instructions and sample words that were used to define the s and sh labels were written with the standard writing system, which is a mix of kanji (Chinese characters), katakana (a Japanese syllabary mainly used to denote foreign words or scientific names), and hiragana (a different Japanese syllabary mainly used for native words). Although all the sample words in Japanese contained word-initial /s/ for the s label or the /ʃ/ sound for the sh label, these word-initial fricatives are not as transparent or as easily decomposed from the Japanese writing system as the English ones, a fact we return to in the discussion.

The presentations of the 2 blocks were counterbalanced within the 19 English listeners and the 20 Japanese listeners. The order of the actual stimuli inside each block was randomized for each individual listener. For each block, listeners responded by pressing a "yes" or "no" button as quickly as they could with the index finger of their dominant hands. A PST (Psychology Software Tools) serial response box was used. Only the accurate data are analyzed here. [See Urberg-Carlson *et al*. (2009) for an analysis of response time data (RTs) from the English-speaking listeners.] One English listener's data turned out to be unusable because of equipment failure and were not included in the analysis, leaving 18 English-speaking listeners.

III. ANALYSES AND RESULTS

A. Logistic regression: Naïve listeners' judgments of children's fricative productions

Because the purpose of the perception experiments was to evaluate the source of cross-linguistic differences in normative data derived from consensus transcriptions of children's productions, the first set of analyses aggregated naïve listeners' judgments in the two language communities. In other words, previously reported error patterns were based on the transcription results where native-speaker transcribers pretend to be naïve in their judgments of children's speech. In our experiments, we used real naïve listeners who did not receive phonetic training to get their judgments of children's fricative productions. In order to be comparable to the previous transcription results, we designed a way to assign each stimulus token a label that is indicative of whether these naïve listeners generally accept that speech sound as /s/ or /ʃ/ in their native languages. More specifically, each stimulus token was labeled as <s>, <sh>, or <neither> based on the following procedure. A token was tagged as <s> if it received yes responses from 70% or more of the listeners within a given language group (70% was the threshold for being significantly different from chance at the $a < 0.05$ level, based on the binomial probability distribution) when the question was "Is this an 's'?". Similarly, a token was labeled as <sh> if it received yes responses from 70% or more of the listeners when the question was "Is this an 'sh'?". Those tokens receiving less than 70% positive responses from all listeners in either block were labeled as <neither>. A breakdown of all the stimuli as classified into different perceived categories in regard to the intended target fricatives is listed in Table II. The table shows that English listeners identified 65% (51 out of 78 tokens) of the intended /s/ productions by English-speaking children/adults to be on-target and 7% (6 out of 78 tokens) to be [ʃ]-for-/s/ substitutions. By contrast, the Japanese listener group identified only 43% (34 out of 78) of the English stimuli as on-target /s/ productions. The two listener groups, however, converge when judging Japanese-speaking children/adults' intended /s/ productions (40% vs 40%). The discrepancy between the judgments of the two listener groups on English intended /s/ tokens (65% vs 43%), but the absence of such a difference for the Japanese stimuli could suggest English-speaking listeners' leniency toward recognizing /s/ in children's speech, or more mature /s/ productions by English-speaking children, or both. Listeners' category judgments, however, reflect indirect inferences of children's speech based on a complex accumulation of acoustic cues in the speech signals. In order to tease production differences apart from perception differences, an analysis probing the relationship between acoustic cues underlying category judgments and the acoustic characteristics of the stimuli was needed.

Logistic regression models were used to analyze the results below category threshold by associating listeners' fricative judgments with specific acoustic cues in the stimuli. The dependent variables were the two perceived categories: <s> (coded as 0) and <sh> (coded as 1). Tokens belonging to the <neither> category were excluded from this analysis and are discussed in detail later. The independent variables were (a) the standardized values of the five spectral acoustic parameters for those tokens that have been identified as either <s> or <sh> by the community, (b) talker language (i.e., whether the stimulus was produced by an English or a Japanese speaker), and (c) the interaction between the standardized values of the five acoustic measures with stimulus language. The reason to include stimulus language together

TABLE III. Results of logistic regression for the two listener groups on the five acoustic parameters as well as on the effect of stimulus language (English vs Japanese). The *p*-values of those predictors that were statistically significant in predicting fricative categories are shown in bold.

	English listeners				Japanese listeners			
Acoustic predictors	Coefficient	Standard error	*Z*-value	*p*-value	Coefficient	Standard error	*Z*-value	*p*-value
M1	−5.4	1.5	3.5	**<0.001**	−5.1	2.0	−2.5	**0.012**
M2	−0.8	0.6	−1.5	0.134	−2.3	1.1	−2.1	**0.032**
M3	−1.3	1.1	−1.2	0.229	−0.6	1.2	−0.6	0.580
M4	−0.01	0.8	−0.02	0.983	−5.3	2.7	−1.9	0.051
Onset F2	1.7	0.7	2.2	**0.026**	4.3	2.0	2.2	**0.027**
Stimulus language	−1.3	1.6	−0.8	0.411	−2.2	1.0	−2.2	**0.029**
M1 × stimulus language	−3.9	3.2	−1.2	0.227	−2.4	3.0	−0.8	0.417
M2 × stimulus language	−2.2	1.5	−1.4	0.149	−0.2	1.5	−0.1	0.912
M3 × stimulus language	−0.9	1.7	−0.5	0.601	−2.4	2.0	−1.2	0.231
M4 × stimulus language	−21.8	8.0	−2.7	**0.006**	−11.7	5.8	−2.0	**0.046**
Onset F2 × stimulus language	0.6	1.1	0.6	0.571	−2.5	2.1	−1.2	0.218

with its interaction with the acoustic predictors as independent variables is that the stimuli from the two languages were mixed into a single block for presentation and tacit awareness of the language from which the fricative came might have influenced the perception of listeners to some extent. Logistic regression allows us to determine the subset of predictors significantly associated with the probability of identifying fricatives. The standardized coefficients of each predictor can then be used to evaluate the relative contributions of different predictors to the overall model. Two logistic regressions were performed, one for each of the two listener groups. Table III shows the results of the logistic regression model for both listener groups.

It is clear from the left part of Table III that English-speaking listeners relied primarily on two acoustic parameters, M1 and onset F2 frequency. The negative coefficient for M1 indicates an association between a lower M1 value and a higher probability of listeners' categorizing a given fricative sound as being /ʃ/. This is exactly in line with our expectations because /ʃ/ has a lower M1 value than /s/. Although the majority of noise in producing /s/ and /ʃ/ is generated when the air stream impinges on the teeth, the difference in spectral mean energy has been attributed primarily to the difference in the front resonating cavity between the two voiceless sibilant fricatives (Stevens, 1998). By the same token, the positive coefficient of onset F2 frequency suggests an increase in probability for the percept of /ʃ/, as the /ʃ/ sound is produced with a constriction further back in the oral cavity, resulting in a higher onset F2 frequency in the vowel spectrum, which is also consistent with expectations. It is also important to note that the absolute value of the coefficient for M1 (5.4) is higher than that of the coefficient for onset F2 frequency (1.7), suggesting a greater predictive power of M1 relative to that of onset F2 in determining fricative categories by English-speaking listeners. In addition, a significant effect of the interaction between M4 and stimulus language was found in the English-speaking listeners' group. This interaction indicates that English-speaking listeners associate M4 in a different way when perceiving their native language as compared with their perception of Japanese stimuli. This interaction term will be discussed again in Sec. III B when probability curves derived for each listener group are described.

The relationship of each predictor to listener perceptions was different for Japanese-speaking listeners, as shown in the right half of Table III. Three acoustic parameters were associated significantly with successful identification of fricative categories. These three parameters were M1, onset F2 frequency, and M2. M1 contributed most to the identification of /s/ and /ʃ/ (its coefficient has the highest absolute value, 5.1, followed by onset F2 frequency with a coefficient with an absolute value of 4.3, and then M2, with a coefficient with an absolute value of 2.3). Similar to the results for the English stimuli, the negative value of the coefficient here indicates that the lower the value of M1, the more likely it was to be judged as /ʃ/. Again, onset F2 frequency was positively correlated with the percept of /ʃ/, as predicted. The third predictor that significantly contributed to the model is M2, which was negatively correlated with the likelihood of perceiving /ʃ/. Because M2 is a measure of the variance of the density distribution of the fricative noise spectrum, the negative coefficient here means that more the compact the spectral shape is (i.e., the lower the M2 value), the more likely the fricative is judged as /ʃ/. This is not surprising, given the fact that the Japanese /s/ sound is described as "less sibilant," which indicates a more diffuse spectral shape than the /ʃ/ sound. In addition to the three acoustic parameters that significantly contributed to the probability of the /s/-/ʃ/ percept, an effect of stimulus language as well as an interaction between stimulus language and M4 were also found to be significant for Japanese-speaking listeners. These effects will be discussed again in Sec. III B.

One thing to note is that M1 and onset F2 frequency are the two primary perceptual correlates of voiceless sibilant fricatives for both listener groups, but they were weighted more similarly by Japanese-speaking listeners (5.1 vs 4.3) than by the English-speaking listeners (5.4 vs 1.7). Figure 2 visually displays the performance of the two listener groups by plotting onset F2 values against those of M1 for all of the English stimulus tokens. It can be observed that the vast majority of the tokens classified as <s> by English listeners have M1 values above 6000 Hz and the great majority of those classified as <sh> have M1 values below 8000 Hz. For onset F2 values, the <sh> tokens occupy a range slightly lower than that of the <s> tokens, although there is overlap between the two categories. A discriminant function line was drawn to

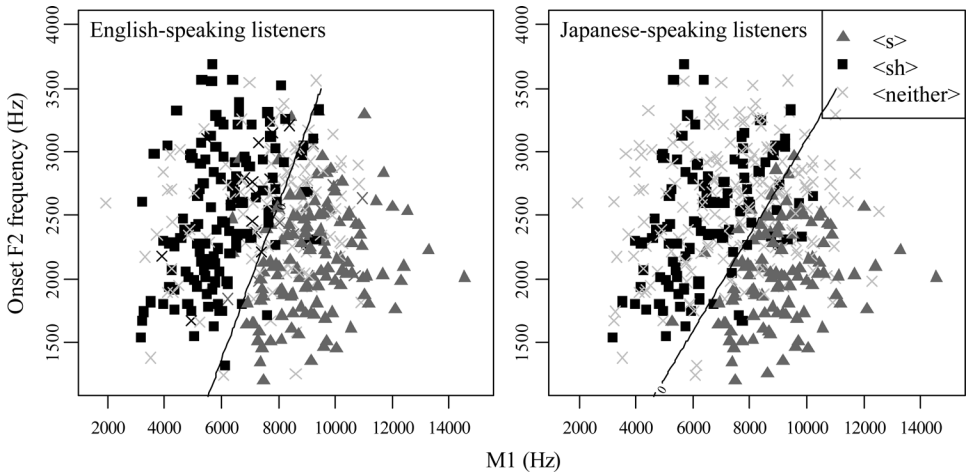

FIG. 2. English-speaking and Japanese-speaking listeners' responses to the stimuli. Black squares represent <s>, naïve native speakers' judgments of a given stimulus being /s/ according to a statistically significant criterion. Gray triangles represent <sh>, naïve native speakers' judgments of a given stimulus being /ʃ/. The crosses are the <neither> cases that did not meet the criterion and thus fall into either the <s> or the <ʃ> category.

help demarcate the boundaries of the two categories. The line is nearly vertical for English-speaking listeners, reflecting the stronger predictive power of M1 relative to onset F2 frequency. By contrast, for Japanese-speaking listeners, greater overlap exists in the M1 dimension between 6000 and 10 000 Hz for the two categories. Furthermore, the overlap in onset F2 values is relatively smaller compared with that for English listeners. As a result, the discriminant function line is shallower for Japanese listeners, reflecting the finding that both M1 and onset F2 contributed relatively equally to the Japanese-speaking listeners' classification of the stimuli.

B. Probability functions and phonemic boundaries

To quantify the phonemic boundaries between the two perceptual categories <s> and <sh>, probability scores transformed from the above logistic regression models were plotted for M1 and onset F2 frequency for English-speaking and Japanese-speaking listeners, as these two are the two primary acoustic parameters shared by both listener groups. These are shown in the upper two panels of Fig. 3. In each of these graphs, acoustic parameter values were arranged from lower to higher, from left to right, along the *x*-axis. The *y*-axis shows the probability scores ranging from 0 to 1, with 0 being "definitely <s>" and 1 being "definitely <sh>." "Phoneme boundary" is defined as the predicted value for a given acoustic parameter when the probability score is equal to 0.5.

In the M1 dimension, both listener groups showed the classical categorical perception pattern (i.e., a sigmoidal identification function). More specifically, the higher the M1 value of a fricative, the more likely listeners were to classify it as <s>; conversely, the lower the M1 value, the more likely

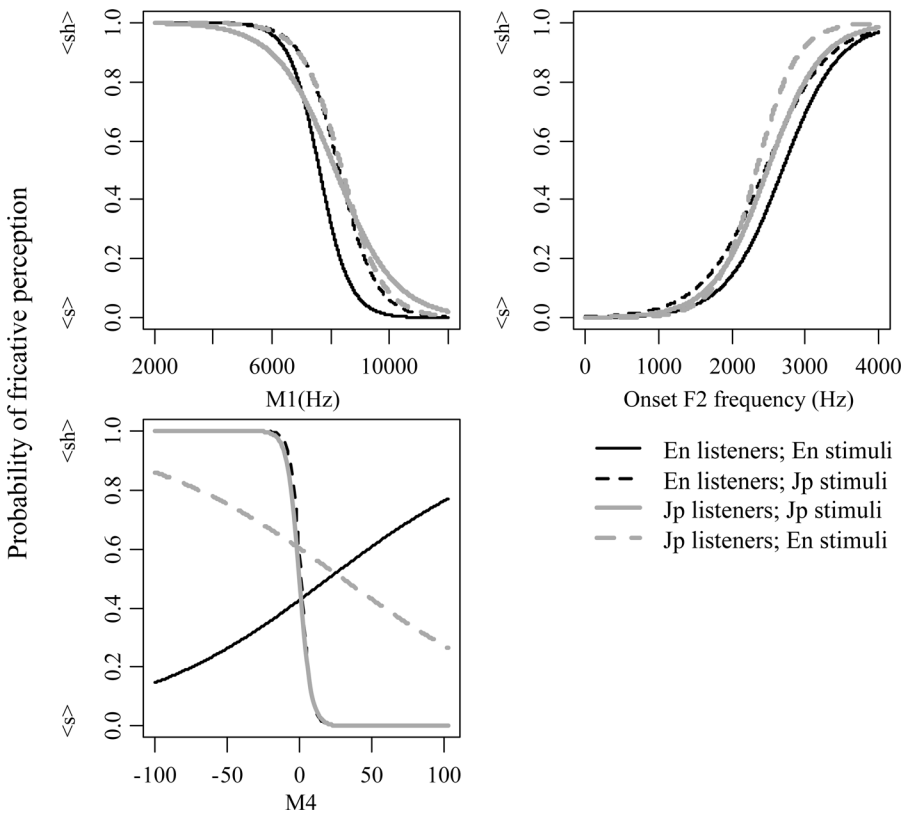

FIG. 3. Probability functions derived from logistic regressions for M1, onset F2 frequency, and M4, respectively. The *y*-axis shows the predicted probability scores of fricative perception, with "1" being 100% <sh> and "0" being 100% <s>. The *x*-axis shows the acoustic values of stimuli in each of the three acoustic dimensions. The black lines describe the predicted English-speaking listeners' responses to English stimuli as a function of acoustic values in M1, onset F2 frequency, or M4; the black dotted lines are English-speaking listeners' responses to Japanese stimuli; the gray lines are Japanese-speaking listeners' perceptions of Japanese stimuli; and the gray dotted lines are Japanese-speaking listeners' perceptions of English stimuli.

listeners were to classify it as <sh>. Japanese-speaking listeners showed shallower slopes, suggesting less-categorical identification, than did English-speaking listeners, especially when judging their native language stimuli. They also have a phoneme boundary approximately 500 Hz higher than that of the English-speaking listeners for <s>. Because M1 is positively correlated with the percept of /s/, a higher phoneme boundary for M1 indicates a smaller range of acceptability for <s> by Japanese-speaking listeners. In the onset F2 dimension, the reverse pattern was found for the probability curves of both groups. This is expected as onset F2 is negatively correlated with the percept of /s/. Therefore, the higher the onset F2 frequency, the less likely it is that a fricative will be judged as <s>. At the same time, when judging native language stimuli in particular, English-speaking listeners showed a higher boundary for the <s> category than Japanese-speaking listeners. Given the negative correlation between onset F2 and the percept of /s/, this higher boundary suggests a larger range of acceptability for <s> by English-speaking listeners.

In addition, a probability function was also described for M4 in order to examine the interaction effects found in the logistic regression models for both English-speaking listeners and Japanese-speaking listeners, as shown in the lower panel of Fig. 3. Both listener groups showed an interaction effect between M4 and stimulus language. It is immediately apparent from the graph that the prediction curves for English-speaking listeners go in different directions for their judgments of native language stimuli and for their judgments of Japanese stimuli. For English-speaking listeners, M4 is positively correlated with the percept of /ʃ/ when listening to fricatives produced by English-speaking children, but negatively correlated with the percept of /ʃ/ when listening to fricatives produced by Japanese-speaking children. Furthermore, the probability curve is very steep for the Japanese stimuli but much shallower for the English stimuli, indicating that M4 has much less predictive power for the latter. In contrast, for Japanese-speaking listeners, the probability functions for the English and Japanese stimuli are in the same direction. Similar to the results for the English-speaking listeners, however, the steepness of the probability functions differs for the two sets of stimuli. The probability curve is very shallow for Japanese-speaking listeners when listening to English stimuli but of perfect sigmoidal shape when listening to Japanese stimuli. This result suggests that both listener groups agreed that M4 is strongly and positively correlated with the percept of /ʃ/ for the Japanese stimuli, whereas the relationship between the percept of /ʃ/ and M4 for the English stimuli is weaker or non-existent.

C. The <neither> cases

It is notable that for both languages some stimuli were not consistently categorized as either <s> or <sh>. In order to investigate the nature of those sounds, the <neither> cases were compared with those identified as either <s> or <sh> using the three acoustic parameters (i.e., M1, M2, and onset F2) that were shown to correlate with listeners' fricative perceptions for English-speaking or Japanese-speaking listeners. Furthermore, a series of *t*-tests was performed to quantify such differences between the <neither> cases and the <s> or <sh> cases in each of the three acoustic

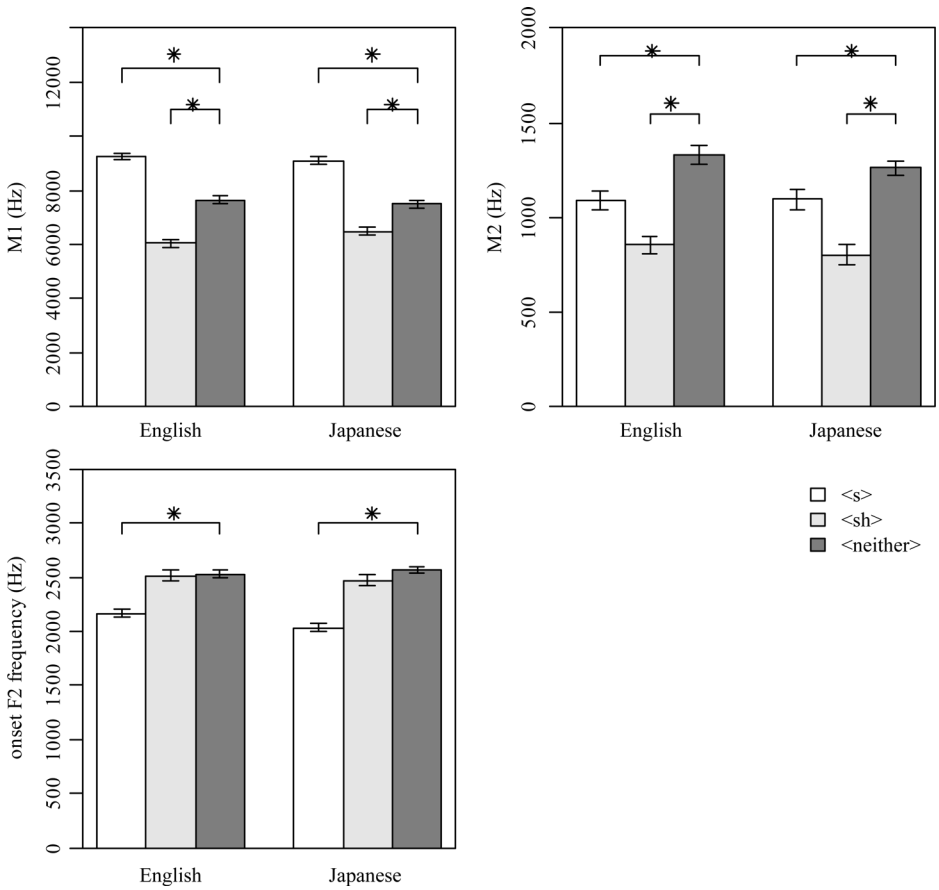

FIG. 4. The mean values of the <s> tokens (the unfilled bars), the <sh> tokens (the light gray bars), and the <neither> tokens (the dark gray bars) for the two listener groups in each of the three acoustic dimensions, respectively. The <s> or the <sh> tokens are defined by receiving more than 70% of yes responses from all the native-speaking listeners when they were asked is this an "s"? or is this an "sh"?, respectively. Error bars indicate one standard error above and below the means. *t*-Tests were performed between the <s>/<sh> tokens and the <neither> tokens. Significantly different means at the level of 0.05 are indicated by an asterisk.

TABLE IV. Results of *t*-tests between the <s>/<sh> tokens and the <neither> tokens for the two listener groups in the three acoustic dimensions. Significant *p*-values at the level of 0.05 are in bold.

Acoustic parameters	Listener groups	Comparison groups	*t*	DOF	*p* value
M1	English-speaking listeners	<s>vs <neither>	7.8864	273	**<0.001**
		<sh>vs <neither>	−7.946	265	**<0.001**
	Japanese-speaking listeners	<s>vs <neither>	7.0887	299	**<0.001**
		<sh>vs <neither>	−4.2029	291	**<0.001**
M2	English-speaking listeners	<s>vs <neither>	−3.5687	273	**<0.001**
		<sh>vs <neither>	−7.1382	265	**<0.001**
	Japanese-speaking listeners	<s>vs <neither>	−2.4075	299	**0.02**
		<sh>vs <neither>	−6.7182	291	**<0.001**
Onset F2	English-speaking listeners	<s>vs <neither>	−6.8466	273	**<0.001**
		<sh>vs <neither>	−0.2205	265	0.08
	Japanese-speaking listeners	<s>vs <neither>	−10.228	299	**<0.001**
		<sh>vs <neither>	−1.7224	291	0.08

dimensions, respectively. The comparison and the results of the *t*-tests are graphically presented in Fig. 4. Specifically, in each of the three acoustic dimensions, the mean values of the three categories (<s>, <sh>, and <neither>) were plotted for the two listener groups separately. Statistically significant comparisons between columns are indicated with an asterisk.

For M1, for both listener groups, those sounds identified as <s> show the highest mean M1 values, whereas those judged as <sh> show the lowest mean values. The <neither> cases have mean values falling into the intermediate range between <s> and <sh>. Four *t*-tests were performed, two for each listener group, between the <neither> cases and the <s> (or <sh>) cases (the *t* statistics are in Table IV). All four comparisons were found to be statistically significant. For M2, the <neither> cases show the highest mean M2 values as compared to either the <s> or the <sh> tokens. Again, all four comparisons were statistically significant, indicating significantly higher M2 values for the <neither> cases relative to the two consistently perceived fricative categories. In the dimension of onset F2 frequency, the <neither> cases again showed higher mean values than the <s> or the <sh> cases. However, only the comparisons between the <neither> and the <s> tokens were found to be statistically significant, while the comparisons between <neither> and <sh> were not. This suggests that the <neither> cases have higher onset F2 values than the <s> cases but share similar onset F2 values with the <sh> category.

These <neither> tokens, therefore, have mean values intermediate between the <s> and <sh> tokens for M1. These tokens also have consistently higher values for M2 than both <s> and <sh>. They also have higher values for onset F2 than <s> but not <sh>. Such acoustic characteristics suggest a more diffuse spectral shape and a less sibilant nature for these tokens. These acoustic properties are consistent with those of nonsibilant fricatives in English such as /f/ or /θ/, as described in Jongman *et al*. (2000), except for the high onset F2 values, which suggests a further back constriction in the oral cavity. It is possible that these tokens were somehow confusable with English nonsibilant fricatives such that it was difficult for native speakers of English to classify them as either <s> or <sh>. For Japanese listeners, such sounds were most likely to be confused with the nonsibilant fricative sound [ç] (which occurs as an allophone of /h/ prior to /i/, as in [çime] "princess") in Japanese.

IV. DISCUSSION

This study has several major findings. First, we observed cross-language differences in adults' perception of children's speech. English-speaking listeners' perceptions of /s/ and /ʃ/ were correlated primarily with M1 and onset F2 frequency, whereas Japanese-speaking listeners' perceptions were correlated with M1, onset F2 frequency, and M2. This finding is compatible with results of a previous study of English-speaking and Japanese-speaking adults' productions of voiceless sibilant fricatives (Li *et al*., 2009), which found that English-speaking adults' /s/ and /ʃ/ productions differ primarily in M1, whereas Japanese-speaking adults distinguish their sibilant fricatives in M1, onset F2, and, marginally, in M2. There are striking parallels between the results of the current perception study and those of the previous production study. In both studies, M1 is the main acoustic parameter that was correlated with both adults' production and perception of voiceless sibilant fricatives for both English speakers and Japanese speakers. Furthermore, Japanese speakers utilize more acoustic dimensions in both producing and perceiving sibilant fricative contrasts than do English speakers. Critically, the current study found evidence that the well-documented asymmetry in the order of acquisition of /s/ and /ʃ/ in English and Japanese may be due to different perceptual norms for adult speakers of these languages. We showed different phoneme boundaries between /s/ and /ʃ/ for both listener groups, based on the probability functions derived from logistic regression models. Particularly, English listeners showed a lower phoneme boundary in the M1 dimension and a higher boundary in the onset F2 dimension than Japanese listeners. Because M1 is positively correlated and onset F2 frequency is negatively correlated with the percept of /s/, these patterns in phoneme boundaries suggest a greater perceptual space for /s/ for English listeners. For Japanese listeners, the opposite pattern was found, with the phoneme boundary between /s/ and /ʃ/ being higher in the M1

dimension and lower in the onset F2 dimension. Their perceptual /s/ space is thus relatively smaller than that of /ʃ/. In other words, when presented with ambiguous or intermediate speech sounds such as those common in children's speech, English listeners are more likely to assimilate them into their /s/ category, whereas Japanese listeners are more likely to assimilate them into their /ʃ/ category. Such a difference in the perceptual range of fricative categories is in accordance with the different acquisition and error patterns in the two languages, where English-speaking children are perceived as correctly producing /s/ earlier and making [s]-for-/ʃ/ substitutions, while Japanese-speaking children are perceived as correctly producing /ʃ/ earlier and making [ʃ]-for-/s/ substitutions.

The fact that Japanese speakers use more acoustic parameters to differentiate the two voiceless sibilant fricatives for both production and perception suggests a less robust phonetic representation of the /s/-/ʃ/ contrast. This may be a reflection of the less robust status of this contrast in the higher-level phonological representation in Japanese. Specifically, while /s/ and /ʃ/ are contrastive in all following vowel contexts in English, Japanese /s/ and /ʃ/ are distinguished only before back vowels. The contrast is traditionally neutralized before front vowels: Only /s/ is permitted before /i/, and only /ʃ/ is permitted before /e/.

The contribution of M2 to the perception of /s/ and /ʃ/ in Japanese may also be related to the specific characteristics of /s/ and /ʃ/ productions in Japanese. M2 describes the variance of the fricative spectrum, which is negatively correlated with the percept of /ʃ/. This suggests a more diffuse spectral shape of /s/ in acoustics and is in accordance with the laminal-dental tongue posture of /s/ in articulation as opposed to the more palatalized posture in producing /ʃ/. The association of M2 with laminality and tongue posture is not novel. For example, Stoel-Gammon *et al.* (1994) compared the American English /t/, which is laminal-dental, with the Swedish /t/, which is an apico-alveolar, in adults' and children's productions and found that M2 is one of the significant parameters of tongue posture that separates the two coronal stops with different articulatory configurations.

We also found that the relative importance of the different acoustic cues differ across the two listener groups. For English-speaking listeners, M1 is a much stronger predictor than onset F2 frequency of sibilant fricative identification. In contrast, Japanese-speaking listeners show a much more similar weighting of M1 and onset F2 in identifying sibilant fricatives. The greater importance of onset F2 frequency to Japanese-speaking listeners may be related to the specific articulatory characteristics of the Japanese /ʃ/. As noted earlier, the production of Japanese /ʃ/ involves a palatalized tongue posture. This effectively shortens the length of the back resonating cavity and thus results in a high onset F2 frequency in the following vowel. In fact, this palatalized posture is so inherently incompatible with low back vowels such as /a/ and /u/ that its transition into the following vowel is characterized by a /j/-like percept owing to coarticulation. Such interpretation is consistent with the results of Toda (2007), where Toda has observed consistently higher onset F2 frequencies across different vowel contexts and across all individual speakers for /ʃ/ than for /s/ produced by Japanese native speakers and concluded that vowel transitions, together with the noise spectra, are equally important components in forming the /s/-/ʃ/ contrast in Japanese.

The result that Japanese listeners rely more on transitional information such as onset F2 frequency may also be explained by the conclusions of Wagner *et al.* (2006). In their study, Wagner *et al.* tested the role of formant transitions in fricative perceptions in five languages: Dutch, German, Spanish, English, and Polish, which differ in their fricative inventories. In a series of experiments, they embedded either natural or conflicting formant transitions in nonsense words containing target /s/ or /f/ and asked the native speakers to identify the target phonemes. They found no effect of formant transitions for /s/ or /f/ in Dutch and German, the two languages that do not have spectrally confusable fricatives present in the native phoneme inventories. Unnatural formant transitions did affect Spanish-speaking listeners' perception of /f/ and English-speaking listeners' perception of /f/ and /s/, as Spanish has a competing fricative /θ/ that is spectrally similar to /f/ and English has both /θ/ and /ʃ/ to compete with /f/ and /s/, respectively. For Polish-speaking listeners, the perception of /s/ relies on transitional information more than that of /f/, because Polish has three other sibilant fricatives (/ʃʲ/, /ɕ/, and /ʂ/) that are spectrally similar to /s/. Our results for Japanese listeners' fricative perception further demonstrate that it is more the presence of any spectrally confusable fricatives than the absolute number of fricatives in the phoneme inventories *per se* that contributes to the increased importance of formant transitions in fricative perception. This is because both English and Japanese share the same number of voiceless sibilant fricatives, and Japanese even has fewer fricatives (four, including /s/, /ʃ/, /ɸ/, and /ç/) compared with English (seven, including /f/, /v/, /s/, /z/, /θ/, /ð/, and /h/) if all fricatives were included, but the Japanese /s/-/ʃ/ contrast is more spectrally similar than the English pair (Li *et al.*, 2009)

One final thing to note is the larger number of the <neither> tokens for Japanese-speaking listeners as compared to English-speaking listeners. We speculate that this difference may be attributable to the Japanese writing system, which mixes phonographic hiragana and katakana with the logographic kanji characters that originated from Chinese and are also used in many of the Chinese languages' orthographies. The hiragana and katakana graphemes are a syllabary, in which each graph or digraph represents a moraic segment. The syllabic nature of the Japanese writing system thus fosters a metalinguistic awareness of syllables more directly than it fosters awareness of individual phonemes. By contrast, the English writing system is alphabetic and fosters awareness of phonemes more directly than it does awareness of syllables. The indirect relationship between the Japanese writing system and phonemes may result in a different representation of the contrast between these two categories in Japanese and English listeners. English listeners may have a more clear-cut categorization between the two sounds because of a writing system that fosters phonemic awareness. Further experiments using tasks that do not rely on listeners' phonemic awareness are needed to identify the degree and the exact cause of Japanese

listeners' perceptual inconsistency. We are actively testing this possibility in our current studies on this topic.

Nevertheless, the most important implication of the current research is the limitation of phonetic transcription in the research of child phonological development. As Edwards and Beckman (2008) noted, transcriptions are traditionally used for two different purposes. One purpose is to apply broad transcription to the evaluation of whether children's speech productions are correct or incorrect as perceived by the immediate speech community. The other purpose is to apply narrow transcription to the description of lower-level phonetic details in children's speech. Edwards and Beckman (2008) argue that these two purposes of transcription are conflicting in nature because the first purpose requires transcribers to categorize children's speech using language-specific perceptual knowledge as if they were naïve listeners, whereas the second purpose requires them to be objective and language neutral. Our current study has demonstrated the existence of such language-specific perceptual strategies that are below the thresholds of category perception in English-speaking and Japanese-speaking adults. We argue that a perception experiment such as ours is a better alternative to achieve the first goal of native-speaker transcription, whereas instrumental analysis is better to accomplish the second goal of the transcription method. In other words, the current study suggests that we cannot simply study children's speech-sound acquisition at the phonological level, assuming a set of universal sound categories in the world's languages and aiming to identify the order of phoneme acquisition in a particular language. Because of the differences in articulatory, acoustic, and perceptual instantiations of what appear to be the same sound category across languages, we need to directly describe children's speech development using methods such as acoustic analysis in combination with native-speaker perception experiments in order to capture the developmental trajectories of child speech as well as to compare them across languages.

ACKNOWLEDGMENTS

Portions of this research were conducted as part of the first author's Ph.D thesis from the Department of Linguistics, Ohio State University, completed in December 2008. This research was supported by NIDCD (National Institute on Deafness and Other Communication Disorders) Grant No. 02932 to J.E., a McKnight presidential fellowship to B.M., and NSF (National Science Foundation) Grant No. BCS0739206 to M.E.B. We are especially grateful to Dr. Mary E. Beckman for her generous contributions and support to the early structure of the study, as well as much valuable advice and many comments to the Ph.D thesis of the first author where this study came from.

Akamatsu, T. (**1997**). *Japanese Phonetics: Theory and Practice* (Lincom Europa, Newcastle), pp. 91–94.

Aoyama, K., Guion, S. G., Flege, J. E., Yamada, T., and Akahane-Yamada, R. (**2008**). "The first years in an L2-speaking environment: A comparison of Japanese children and adults learning American English," IRAL **46**, 61–90.

Beckman, M. E., Yoneyama, K., and Edwards, J. (**2003**). "Language-specific and language universal aspects of lingual obstruent productions in Japanese-acquiring children," J. Phonetic Soc. Japan **7**, 18–28.

Behrens, S. J., and Blumstein, S. E. (**1988**). "Acoustic characteristics of English voiceless fricatives: A descriptive analysis," J. Phonetics **16**, 295–298.

Best, C. T. (**1990**). "Adult perception of nonnative contrasts differing in assimilation to native phonological categories (A)," J. Acoust. Soc. Am. **88**, S177–S178.

Best, C. T. (**1995**). "A direct realist perspective on cross-language speech perception," in *Cross-Language Speech Perception*, edited by W. Strange and J. J. Jenkins (York Press, Timonium, MD), pp. 171–204.

Best, C. T., and McRoberts, G. W. (**2003**). "Infant perception of non-native contrasts that adults assimilate in different ways," Lang. Speech **46**, 183–216.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Second Language Speech Learning*, edited by M. J. Munro and O.-S. Bohn (John Benjamins, Amsterdam, The Netherlands), pp. 13–34.

Boersma, P., and Weenink, D. (**2005**). PRAAT: Doing phonetics by computer (version 5.0.24) [Computer program]. Retrieved April 17, 2005, from http://www.praat.org

Edwards, J., and Beckman, M. E. (**2008**). "Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes," Lang. Learn. Dev. **4**(1), 122–156.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, The Netherlands), pp. 169–185.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**(1), 115–123.

Fox, R. A., and Nissen, S. L. (**2005**). Sex-related acoustic changes in voiceless English fricatives. J. Speech Lang. Hear. Res. **48**, 753–765.

Funatsu, S. (**1995**). Cross language study of perception of dental fricatives in Japanese and Russian, in *Proceedings of the XIIIth International Congress of Phonetic Sciences (ICPhS '95)*, Vol. 4, edited by K. Elenius and P. Branderud (KTH and Stockholm University, Stockholm, Sweden), pp. 124–127.

Gibbon, F., Hardcastle, W. J., and Dent, H. (**1995**). "A study of obstruent sounds in school-age children with speech disorders using electropalatography," Eur. J. Disord. Comm. **30**, 213–225.

Halle, M., and Stevens, K. N. (**1997**). "The postalveloar fricatives of Polish," in *Speech Production and Language: In Honor of Osamu Fujimura*, Vol. 13, edited by Hajime Hirose and Hiroya Fujisaki Shigeru Kiritani (Mouton de Gruyter, Berlin), pp. 176–191.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang. Speech **1**, 1–7.

Hirai, S., Yasu, K., Arai, T., and Iitaka, K. (**2005**). "Perceptual weighting of syllable-initial fricatives for native Japanese adults and for children with persistent developmental articulation disorders," Sophia Linguist. **53**, 49–76.

Hughes, G. W., and Halle, M. (**1956**). "Spectral properties of fricative consonants," J. Acoust. Soc. Am. **28**, 303–310.

Iverson, P., and Kuhl, P. (**1995**). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," J. Acoust. Soc. Am. **97**, 553–562.

Jakobson, R. (**1941/1960**). *Child Language, Aphasia, and Phonological Universal* (Mouton, The Hague, The Netherlands), pp. 47–57.

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

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford, UK), pp. 145–164.

LaRiviere, C., Winitz, H., and Herriman, E. (**1975**). "The distribution of perceptual cues in English prevocalic fricatives," J. Speech Hear. Res. **18**, 613–622.

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

Locke, J. L. (**1983**). *Phonological Acquisition and Change* (Academic Press, New York, NY), pp. 64–65.

Miccio, A. W., Forrest, K., and Elbert, M. (**1996**). "Spectra of voiceless fricatives produced by children with normal and disordered phonologies," in

Pathologies of Speech and Language: Contributions of Clinical Linguistics and Phonetics, edited by T. Powell (ICPLA, New Orleans, LA), pp. 223–236.
Nakata, K. (**1960**). "Synthesis and perception of Japanese fricative sounds," J. Radio Res. Lab. **7**(2), 319–333.
Narayanan, S. S., Alwan, A. A., and Haker, K. (**1995**). "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am. **98**(3), 1325–1347.
Nissen, S. L., and Fox, R. A. (**2005**). "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," J. Acoust. Soc. Am. **118**(4), 2570–2578.
Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**(3), 351–382.
Nittrouer, S. (**1995**). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," J. Acoust. Soc. Am. **97**(1), 520–530.
Nittrouer, S. (**1996**). "Discriminability and perceptual weighting of some acoustic cues to speech perception by three-year-olds," J. Speech Hear. Res. **39**, 278–297.
Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**(2), 711–719.
Nittrouer, S., and Lowenstein, J. H. (**2010**). "Learning to perceptually organize speech signals in native fashion." J. Acoust. Soc. Am. **127**, 1624–1635.
Nittrouer, S., and Miller, M. E. (**1997**). "Developmental weighting shifts for noise components of fricative-vowel syllables," J. Acoust. Soc. Am. **102**(1), 572–580.
Pierre A., and Best, C. T. (**2007**). "Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters," J. Acoust. Soc. Am. **121**, 2899–2914.
Sander, E. K. (**1972**). "When are speech sounds learned?" J. Speech Hear. Disord. **37**, 55–63.
Scobbie, J. M. (**1998**) "Interactions between the acquisition of phonetics and phonology." In Papers from the 34th Annual Regional Meeting of the Chicago Linguistic Society, Volume II: The Panels, edited by M. C. Gruber, D. Higgins, K. Olson, and T. Wysocki (Chicago Linguistics Society, Chicago), pp. 343–358.
Scobbie, J. M., Gibbon, F., Hardcastle, W. J., and Fletcher, P. (**2000**). "Covert contrast as a stage in the acquisition of phonetics and phonology," in *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*, edited by M. Broe and J. Pierrehumbert (Cambridge University Press, Cambridge), pp. 194–203.
Shadle, C. H. (**1991**). "The effect of geometry on source mechanisms of fricative consonants," J. Phonetics **19**(3–4), 409–424.
Shadle, C. H., and Mair, S. J. (**1996**). "Quantifying spectral characteristics of fricatives," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP 96)*, Philadelphia, pp. 1517–1520.
Smit, A. B., Hand, L., Frieilinger, J. J., Bernthal, J. E., and Bird, A. (**1990**). "The Iowa articulation norms project and its Nebraska replication," J. Speech Hear. Dis. **55**, 29–36.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge), pp. 379–388.
Stevens, K. N., Li, Z., Lee, C., and Keyser, S. J. (**2004**). "A note on Mandarin fricatives and enhancement," in *From Traditional Phonology to Modern Speech Processing*, edited by H. Fujisaki, G. Fant, J. Cao, and Y. Xu (Foreign language teaching and research press, Beijing), pp. 393–403.
Stoel-Gammon, C., Williams, K., and Buder, E. (**1994**). "Cross-language differences in phonological acquisition: Swedish and American /t/," Phonetica **51**, 146–158.
Templin, M. (**1957**). *Certain Language Skills in Children*, Vol. 26 (University of Minnesota, Minneapolis), pp. 19–60.
Toda, M. (**2007**). "Speaker Normalization of fricative noise: Considerations on language-specific contrast," in *Proceedings of the XVI International Congress of Phonetic Sciences*, Saarbriicken, Germany, pp. 825–828, www.icphs2007.de.
Toda, M., and Honda, K. (**2003**). "An MRI-based cross-linguistic study of sibilant fricatives," in *Paper Presented at the 6th International Seminar on Speech Production*, Manly, Australia.
Urberg-Carlson, K., Munson, B., and Kaiser, E. (**2009**). "Gradient measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness," J. Acoust. Soc. Am. **125**, 2529.
Wagner, A., Ernestus, M., and Cutler, A. (**2006**). "Formant transitions in fricative identification: The role of native fricative inventory," J. Acoust. Soc. Am. **120**(4), 2267–2277.
Wellman, B., Case, I., Mengert, I., and Bradbury, D. (**1931**). "Speech sounds of young children," Univ. Iowa Stud. Child Welfare **5**, 1–82.
Werker, J. F., and Lalonde, C. E. (**1988**). "Cross-language speech perception: Initial capabilities and developmental change," Dev. Psychol. **24**(5), 672–683.
Werker, J. F., Cohen, L. B., Lloyd, V., Casasola, M., and Stager, C. L. (**1998**). "Acquisition of word-object associations by 14-month-old infants," Dev. Psychol. **34**(6), 1289–1309.
Whalen, D. H. (**1984**). "Sub categorical phonetic mismatches slow phonetic judgments," Percept. Psychophys. **35**, 49–64.
Whalen, D. H. (**1991**). "Perception of English /s/–/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices," J. Acoust. Soc. Am. **90**(4), 1776–1785.
Yasuda, A. (**1970**). "Articulatory skills in three-year-old children," Stud. Phonol. **5**, 52–71.