CHILD DEVELOPMENT

Child Development, July/August 2012, Volume 83, Number 4, Pages 1303–1315

Language-Specific Developmental Differences in Speech Production: A Cross-Language Acoustic Study

Fangfang Li
University of Lethbridge

Speech productions of 40 English- and 40 Japanese-speaking children (aged 2–5) were examined and compared with the speech produced by 20 adult speakers (10 speakers per language). Participants were recorded while repeating words that began with ''s'' and ''sh'' sounds. Clear language-specific patterns in adults' speech were found, with English speakers differentiating ''s'' and ''sh'' in 1 acoustic dimension (i.e., spectral mean) and Japanese speakers differentiating the 2 categories in 3 acoustic dimensions (i.e., spectral mean, standard deviation, and onset F2 frequency). For both language groups, children's speech exhibited a gradual change from an early undifferentiated form to later differentiated categories. The separation processes, however, only occur in those acoustic dimensions used by adults in the corresponding languages.

Language development is a process affected by both innate factors and environmental influences. When compared across cultures, children show universal patterns in their development of articulatory motor control. For example, within the 1st year of life, children from every culture undergo similar stages, beginning with vocalizing vowel-like utterances (i.e., cooing), and progressing to combined consonant–vowel sequences (i.e., babbling). Further, children ubiquitously produce vowels and stop consonants first, followed by fricatives and liquids (Jakobson, 1941/1960; Locke, 1983). These developmental commonalities suggest some shared biological constraints that structure the process of sound acquisition, regardless of the linguistic environment (Kent, 1992).

Nevertheless, exposure to a specific language can also exert profound effects on the way children develop speech. In a cross-language study of infant babbling (Boysson-Bardies, Halle, Sagart, & Durand, 1989), vowel productions by 10-month-old infants from English, French, Arabic, and Cantonese backgrounds were spectrally analyzed. The results revealed language-specific patterns in infant vowel articulations that parallel those of adults. Similar language-specific acoustic differences were found in the production of the consonant /t/ by 30-month-olds speaking American English or Swedish (Stoel-Gammon, Williams, & Buder, 1994).

Shared sounds across languages provide a natural venue for the investigation of how biological and environmental factors interact in the course of phonological development. In particular, examining how children acquire analogous sounds in different languages will address the question of whether there exists a universal sound acquisition sequence or not, a question that has instigated extensive debate since the last century. For instance, both English and Japanese have the consonant /s/ (as in *sea*) and /ʃ/ (as in *sheep*) in their phoneme inventories. According to Jakobson (1941/1960) and Locke (1983), speech sounds produced toward the front of the oral cavity (e.g., /s/) will be mastered prior to those articulated further back in the oral cavity (e.g., /ʃ/), regardless of the ambient language that children are exposed to. However, large-scale normative studies conducted in the United States and in Japan documented an earlier emergence of /s/ relative to /ʃ/ for English-speaking children (Smit, Hand, Frieilinger, Bernthal, & Bird, 1990; Templin, 1957), but the opposite pattern for Japanese-speaking children (Nakanishi, Owada, & Fujita, 1972; Yasuda, 1970).

Portions of this research were conducted as part of the author's Ph.D. thesis from the Ohio State University, Department of Linguistics, completed in December 2008. This research was supported by NIDCD Grant 02932 to Dr. Jan Edwards. I would like to express my appreciation to Dr. Jan Edwards and to my advisor, Dr. Mary E. Beckman, for their generous support in data collection and valuable advice on data analysis. I thank the children who participated in the task, the parents who gave their consents, and the schools at which the data were collected. I also thank the editor, the two anonymous reviewers, as well as Dr. Jennifer Mather for their helpful comments. Finally, I would like to thank Dr. Tamiko Ogura for her generosity in sharing unpublished norming data on Japanese CDI with me.

Correspondence concerning this article should be addressed to Fangfang Li, Department of Psychology, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, Canada T1K 3M4. Electronic mail may be sent to fangfang.li@uleth.ca.

 0009-3920/2012/8304-0015
DOI: 10.1111/j.1467-8624.2012.01773.x

The debate over the existence of a universal chronology in speech sound acquisition is exacerbated by the traditional phonetic transcription method used by the above-mentioned studies. The issue with the phonetic transcription method lies in its dependence on adult auditory impressions and the use of International Phonetic Alphabet (IPA) symbols for describing children's speech. Particularly for cross-language research, the drawback to using the phonetic transcription method is clear: The transcription system interprets children's productions through a language-specific filter of adult sound categories and therefore is likely to distort and mischaracterize the true speech production patterns of children (Edwards & Beckman, 2008; Kent, 1996; Scobbie, 1998). Thus, objective methods of assessing speech production, such as acoustics, are needed to compare children's articulation across languages, free from adult perceptual biases.

Another related controversy surrounds the question of whether children acquire one sound after another in a discrete fashion or if they incrementally approximate adult-like forms in a continuous manner. Most studies that employ the transcription method describe the process of phonological development as mastering one sound after another; however, this pattern of sounds emerging fully formed and discrete from one another may be an artifact of the auditory-based transcription method. Adults are biased to perceive speech categorically, resulting in greater perceived differences between speech signals across categorical boundaries than within them (Liberman, Harris, Hoffman, & Griffith, 1957). This nature of speech perception is likely to cause adult listeners to rigidly categorize intermediate productions of children's speech by superimposing subjective interpretations on these immature speech attempts. An alternative theoretical position states that phonemes emerge gradually as children learn to coordinate different aspects of their vocal physiology and fine-tune their articulatory precision (MacNeilage & Davis, 1990; Nittrouer, 1993). This continuous view of speech development is mostly supported by instrumental studies of children's speech (Kent, 1976; Macken & Barton, 1980; Nittrouer, 1995).

The Current Study

The current study aims to tackle the two theoretical problems in child speech development—(a): universality versus linguistic specificity (b) discreteness versus continuity—by comparing the acoustics of /s/ and /ʃ/ sounds produced by English- and Japanese-speaking children. These two sounds are interesting because they are shared by the two languages, but were suggested to developmentally emerge in an opposite order by previous transcription literature. Acoustic analysis techniques will help to determine whether such cross-language differences do indeed exist in children speaking the two languages, or the reported difference in acquisition order is actually due to perceptual distortions introduced through the transcription process. A common early sound form may reflect natural propensities in the development of motor control that are universal across languages, whereas a different early form of an analogous sound would suggest a greater environmental influence. In regard to the second theoretical question of discreteness versus continuity in development, I predicted a continuous acquisition pattern in both language groups based on previous research using instrumental methods.

In order to determine language-specific phonetic differences in the linguistic environment that children are exposed to, adult productions of the two sounds were examined and compared between the two language groups. To identify age-related developmental patterns, children's productions were analyzed across age groups within each language. These patterns were then analyzed across languages for similarities and differences. The following section describes the relevant acoustic measures for the articulation and perception of the two sounds in English and Japanese, which were then applied to describe children's speech.

The Articulations and Acoustics of /s/ and /ʃ/ in English and Japanese

Both /s/ and /ʃ/ are created by air passing through narrow constrictions formed by the tongue and the roof of the mouth, resulting in turbulent noise (Fant, 1960; Ladefoged & Maddieson, 1996). In English, /s/ is produced with the tongue apex/blade raised toward the middle of the alveolar ridge, just behind the upper incisors (Ladefoged & Maddieson, 1996). The /ʃ/ sound is articulated in a similar manner except with a more posterior location of the lingual constriction (Narayanan, Alwan, & Haker, 1995). In Japanese, /s/ has been reported to be articulatorily similar but perceptually less sibilant than the English /s/. The Japanese /ʃ/, different from the English /ʃ/, involves an elevation of both the tongue blade and the tongue body toward the hard palate, creating a relatively long and flat constriction (Akamatsu, 1997). Despite these articulatory differences, however, the two ''back'' fricatives are readily assimilated into each other for second language

learners. For example, Japanese /sɯʃi/ is perceived as /suʃi/ in English and English /ʃak/ is perceived as /ʃakkɯ/ in Japanese.

A widely used analysis method for fricatives is called spectral moments analysis. Spectral moments analyses treat the fricative noise spectrum as a probability density distribution and quantify its shape characteristics through moments (Forrest, Weismer, Milenkovic, & Dougall, 1988). For voiceless sibilant fricatives such as /s/ and /ʃ/ in English and Japanese, the first moment (M1) and the second moment (M2) are most relevant. M1 calculates the mean frequency of a fricative noise spectrum and is inversely related with the length of oral cavity in front of the point of constriction (Jongman, Wayland, & Wong, 2000; Shadle, 1991). The M1 value for /s/ is expected to be higher than that of /ʃ/ because /s/ is articulated with a more anterior lingual constriction than /ʃ/, which results in a shorter front resonance cavity. The second moment, M2, calculates the standard deviation of a fricative noise spectrum. M2 is particularly useful in describing the diffuse spectral shape of Japanese /s/ resulting from the lower degree of sibilance (Li, Edwards, & Beckman, 2009). The M2 value of the Japanese /s/ is expected to be higher than that of English /s/ or Japanese /ʃ/, both of which are more perceptually sibilant.

For Japanese, onset F2 frequency is another acoustic parameter that has been shown critical in differentiating fricative contrasts acoustically (Funatsu, 1995; Halle & Stevens, 1997). Onset F2 frequency refers to the second formant frequency taken at the onset of the vowel, immediately following a fricative. This transitional parameter cues the length of the vocal tract cavity behind the oral constriction and has been shown to correlate negatively with the length of the back cavity (Halle & Stevens, 1997). In Japanese, the constriction for /ʃ/ is longer than that of /s/ due to the elevation of the tongue blade and body during articulation. This effectively shortens the length of the back cavity and, as a result, /ʃ/ has a characteristically high onset F2 frequency value as compared to that of /s/ (Toda, 2007). For the current study, all three acoustic measures, M1, M2, and onset F2 frequency, were employed to analyze adult productions. These same acoustic parameters were subsequently used to analyze children's speech productions.

Method

Participants

For each language group, 10 adult participants aged from 18 to 30 (gender balanced) and 40 children aged 2–5 (gender balanced) were recruited in their home country. Specifically, all English-speaking adults were recruited from Columbus, Ohio, and all Japanese-speaking adults were recruited from Tokyo, Japan. All speakers tested had normal hearing and had passed a hearing screening test using otoacoustic emissions at 2000, 3000, 4000, and 5000 Hz. No adult participants reported histories of speech, language, or hearing problems. No children tested had speech, language or hearing problems, according to reports from teachers or parents.

Task and Materials

In both countries, children were tested individually in a quiet room in a day-care center or preschool. They were engaged in a word-repetition task that was facilitated through a computer program called ''show and play.'' At the beginning of the task, each child was seated in front of a computer monitor. On screen, children were presented with a display that had a ladder along the left margin with a duck at the bottom of it and a picture to the right of it. Each child was told to play a computer game that involved talking to a duck on the screen. The goal of the game was to help the duck go all the way up to the top of the ladder by repeating to it some words that the computer said first. They were told that each time they repeated a word, the duck would climb one step up and that they won the game if the duck reached the top of the ladder. Each child was provided with a few words to practice before entering into the testing phase. It was deemed as successful whenever each child made an attempt to repeat a word following the audio prompt, regardless of whether the attempt contained articulation errors or not. Following a repeat attempt, each child was then directed to the next word. Each child was usually not given a second try on the same word unless: (a) no audible vocalizations was present in the first attempt, (b) the first attempt overlapped significantly with background noise such as a door slamming, or (c) the child produced a completely different word based on naming the picture (e.g., a child said ''chair'' when prompted with ''sofa''). Children usually found this task entertaining and easy to complete. The adult task was similar in that they were also asked to repeat the word given by the prompt. They were also presented a screen with a duck, but were told that the task was mainly designed for young children. The same rules used for the children to determine whether a second attempt was allowed were also applied for the adults.

The audio prompts were prepared and recorded by having a native speaker phonetician use child-directed speech and say a list of words by repeating each word five times in succession. The list of words contained word-initial fricatives as well as stops and affricates such as /t/ and /tʃ/. Individual trials for each word were extracted from the full sound recording and were edited in such a way that a new list of words was created with a randomized order of all the target words. Five naïve speakers were each placed in a sound attenuated booth and then asked to listen to the new random order recording and repeat what they heard. Their productions were recorded and subsequently transcribed by a trained phonetician. Only the trials that were accurately repeated by all five naïve listeners were selected for audio prompts. The visual prompts used in this study consisted of cultural-appropriate pictures of objects commonly known to children in each country.

As discussed in Edwards and Beckman (2008), the word-repetition method offers several advantages for cross-language research with young children. One advantage is that it ensures elicitation of the target consonant sounds in comparable vowel contexts across children speaking different languages. It is also easy to identify the sound target when children make articulation errors. Another advantage is that the word-repetition method is lower in task demand when compared to a picture naming task. Furthermore, Edwards and Beckman suggest that the repetition method does not have the ''confound of age effect, with more words being produced spontaneously in response to pictures by older children who have larger vocabularies, and more words being produced as imitations of a subsequent verbal prompt by younger children who have smaller vocabularies'' (p. 4).

The sound stimuli used in this study consisted of words beginning with fricative-vowel sequences selected based on the following principles. First, the vocalic contexts following target fricatives were matched as closely as possible. English has twice the number of monophtong vowels than Japanese (10 vs. 5). For English, vowels that have similar coarticulatory effects were roughly grouped into a set of five categories *i*, *e*, *a*, *o*, *u*, the only monophthong vowels that Japanese has. Specifically, both /i/ and /ɪ/ were included in the *i* category, all three low back vowels /a/, /ɔ/, /ʌ/ were included in the *a* category, and the vowels /ɛ/ and /e/ were collapsed into the *e* category. There were three target words for each consonant–vowel (CV) sequence, except in cases where some CV sequences could not be elicited owing to phonotactic constraints; for example, in Japanese, /s/ cannot be followed by /i/ and is rarely followed by /e/. Second, only words that were picturable were selected in order to ensure maximal informativeness of the visual prompts. Third, words that are familiar to young children were selected over unfamiliar words when possible; however, word familiarity had to be compromised in some cases to match vowel contexts or to ensure word picturability. More than half of the words selected are those that at least some 2-year-olds are able to produce, according to the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 2000) in English and the adapted version in Japanese (JCDI; T. Ogura, personal communication, March 9, 2011).

Equipment

During the word-repetition task, audio prompts were played through a set of portable speakers attached to the sides of the computer screen. All participants were recorded using Marantz PMD660 professional portable flash card recorders and AKG C5900 microphones. Participants were seated in front of an IBM laptop computer, with the microphone secured on a microphone stand approximately 20 cm away from the participant's mouth. The recordings were made using a 44.1 kHz sample rate and at a 16-bit quantization.

Acoustic Analysis Procedure

The program Praat (Boersma & Weenink, 2005) was used to segment words from larger sound files and was used to identify fricative boundaries. The beginning of a fricative is defined both by a clear increase in the frication noise amplitude in the waveform and by the occurrence of white noise in a frequency band above 2000 Hz in the spectrogram. The end of the fricative is defined as the beginning of the following vowel and the first zero crossing of an upswing pitch cycle of the first periodic glottal pulse of the vowel.

For the spectral moments analysis, a fast Fourier transform (FFT) spectrum was extracted over a 40 ms Hamming window centered around the midpoint of the fricative noise. The middle 40 ms window was chosen because it is the steadiest portion of the fricative noise and is least likely to be influenced by amplitude build-up at the beginning of the fricative or the transitional change into the following vowel. Over the middle-40-ms-window slice, the two spectral parameters M1 and M2 were

calculated; M1 is the mean frequency and M2 is the standard deviation of the frequency distribution in the spectrum.

Onset F2 frequency was measured at the beginning of the vowel that immediately followed the target fricative. As the name indicates, onset F2 frequency measures the second formant of the vowel at onset. (Please refer to Figure 1 for illustration of the extractions of the three acoustic parameters.) The setting used in Praat to estimate onset F2 frequency was an LPC analysis specified for five formants (10 coefficients) calculated over a range from 0 to 5500 Hz for adults and a range from 0 to 7000 Hz for children. The window length was 0.025 ms. All calculations were made without pre-emphasis, because the technique of pre-emphasis gains higher-frequency amplitude at the expense of suppressing the lower frequency amplitude, and the major identifiable differences between /s/ and /ʃ/ are located in the relatively lower frequency range (Tabain & Watson, 1996).

For English-speaking participants, 98% of all the /s/-target trials and 99% of all /ʃ/-target trials were successfully elicited using the word-repetition task. For Japanese subjects, the success rate for eliciting the /s/ target and the /ʃ/ target was 98% and 99%, respectively. Further, responses with clear burst-like interruptions at the beginning or within the frication were excluded. The burst-like interruptions usually results in percepts of stops or affricates and would render the acoustic analysis paradigm designed for fricatives impracticable. Fricative-like responses with less than a 40-ms duration were also excluded. For English subjects, 86% of the successful trials were analyzed for the /s/ target and 91% for the /ʃ/ target. For Japanese subjects, 83% of the successful repetitions were analyzed for the /s/ target and 86% were analyzed for the /ʃ/ target. The total responses analyzed consisted of 533 responses for adults and 1,628 responses for children.

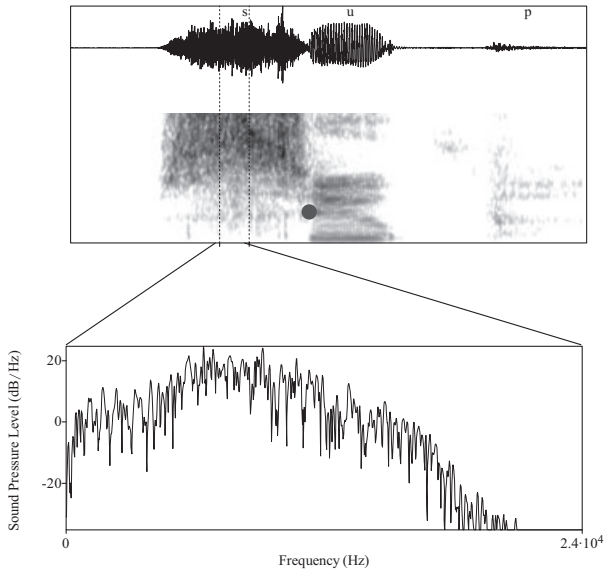

Figure 1. Illustration of the calculation procedures using acoustic analysis software, Praat, for the three acoustic parameters for /s/ in word *soup*.

Note. M1 and M2 calculate the mean and the standard deviation of noise spectrum (the lower panel) extracted from the middle-40-ms window in the sound wave (the upper panel). Onset F2 frequency measures the second formant frequency of the vowel immediately following the target fricative, as indicated by the dot at 1948 Hz in the spectrogram (the middle panel).

Results

Adult Speech Patterns

Spectral differences between the two fricative categories and between the two language groups were compared using the three acoustic parameters M1, M2, and onset F2 frequency. It was necessary to separate the two genders as fricative spectra may differ as a function of gender because females have shorter vocal tracts. For sibilants, the resonating cavity in front of the lingual constriction for female speakers is expected to be shorter than that for males in proportion to the length of the whole vocal tract, thus yielding a higher spectral peak in the spectrum of productions for female speakers relative to male speakers (Fuchs & Toda, 2010; Jesus & Shadle, 2002; Jongman et al., 2000). Figure 2 describes the adult speech patterns based on the productions of the five females only, as adult males' productions are analogous to the females' patterns, but in different absolute acoustic values and therefore will not be reported separately; however, productions of both genders were included in the statistical analysis in the next section. In Figure 2, histograms of the raw counts of fricative tokens were plotted for the three acoustic parameters (M1, M2, onset F2). For the convenience of comparison between the two languages, histograms of English-speaking adults and those for Japanese-speaking adults were arranged side by side for each acoustic parameter. For M1, English speakers showed a clear difference in the distribution between the two fricative targets. Most of the target /s/ sounds had M1 values above 6000 Hz, whereas those of most of the target /ʃ/ tokens were below 6000 Hz. In contrast, a certain degree of overlap exists in the M1 dimension between the two target fricatives for Japanese-speaking

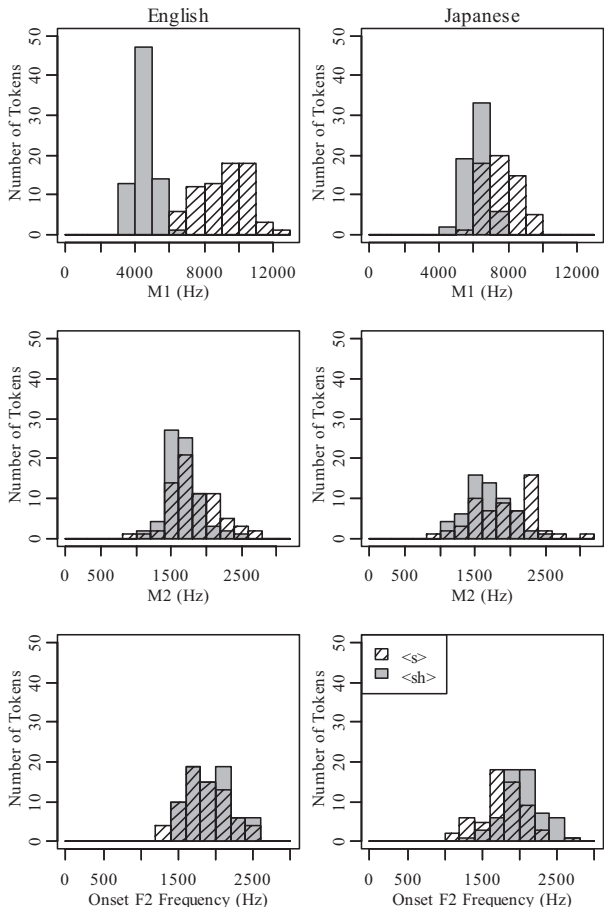

Figure 2. Histograms of token frequency distributions of the two target fricatives in the three acoustic dimensions: M1, M2 and onset F2 frequency, in the adult productions (females only) for each of the two languages.

adults. The partial separation between Japanese /s/ and /ʃ/ in the M1 dimension suggests insufficient power of M1 in differentiating the two categories. In the M2 dimension, for both language groups, considerable overlap exists between the two target fricatives over the range between 500 and 3000 Hz. Despite overlapping, the sound /s/ generally has a higher M2 value than /ʃ/, suggesting a more diffuse spectral shape of /s/. As for onset F2 frequency, both language groups show overlap between the two targets, with the values of /s/ generally lower than those for /ʃ/. However, the level of overlap for the Japanese-speaking adults is relatively less than the level of overlap for the English-speaking adults; this outcome suggests that onset F2 can distinguish the /s/–/ʃ/ distinction for Japanese better than for English.

The relevance and importance of each acoustic parameter in distinguishing /s/ and /ʃ/ in adult speech was formally tested using multivariate logistic regressions called mixed effects logit models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). A mixed logit model is a type of generalized linear mixed effects model that has a linkage of logit function (Jaeger, 2008). The logit function is ideal for binomial outcomes such as in the case of comparing /s/ and /ʃ/ in English and Japanese. The mixed effects model allows for evaluation of the effects of interest (fixed effect) and those of manipulations such as subjects and items (random effects). Unlike regular logistic regression, the generalized linear mixed effects model has the advantage of dealing with variability from clustering data across individual speakers or items that randomly fluctuate. The mixed effects model is also called the hierarchical linear regression model since it admits covariance nested inside subjects as well as other grouping factors that are organized hierarchically. In the current study, two mixed effects logit models were applied using R (R Development Core Team, 2008), each on one language. For each model, the dependent variables are */s/* (coded as *0*) and */ʃ/* (coded as *1*) and the independent fixed effects variables are *z*-score normalized values for each of the three acoustic parameters. The independent random effect variable is individual speakers (10 speakers for each language including both male and female speakers).

The results of the mixed effect logit model for English speakers demonstrated that M1 alone is able to account for approximately 90% of the variability; this outcome makes the fitting of the generalized linear mixed effects model impossible. In other words, the difference in M1 between the two fricative categories is large enough for M1 to predict fricative category by itself in English. For Japanese speakers, however, all three parameters were significant in the model and the results of model fitting are displayed in Table 1. The estimated coefficient of each parameter indicates their relative importance to the model because all the acoustic values are normalized prior to data entry. The nearly equal values of the estimated coefficients of the three parameters suggest an equal amount of predictive power attributable to fricative identity for Japanese speakers. Also, the direction (positive vs. negative) of the estimated coefficient for each parameter suggests the direction of association between acoustic value of individual parameter and the likelihood of the /ʃ/ category (because /ʃ/ is coded as 1 in the model). Specifically, the negative coefficient for M1 in the results shows the negative correlation between M1 and /ʃ/ likeness. This result is consistent with the observation that the energy of the /ʃ/ sound occupies a lower frequency range relative to that of /s/,

Table 1
Results From Mixed Effects Logistic Regression Model on Japanese Adult Productions of /s/ and /ʃ/

	Estimate	Standard error	*z*-value	*p* value
Intercept	23.172	4.486	5.165	<.001
M1	−0.003	0.001	−6.213	<.001
M2	−0.003	0.001	−3.072	.002
Onset F2 frequency	0.003	0.001	2.845	.004

Note. The dependent variables are the two target fricative categories (/s/ vs. /ʃ/) and the independent variables are *z*-normalized values of the three acoustic parameters (fixed effect) and individual speakers (random effect).

as presented in Figure 2. M2 is also shown to correlate negatively with the probability of /ʃ/ and suggests a more diffuse spectral shape for /s/ than for /ʃ/. Onset F2 frequency, on the other hand, is positively associated with the likelihood of /ʃ/. This is because of the shorter length of the back cavity in producing /ʃ/ due to both the more posterior constriction placement and elevated tongue posture in Japanese.

Children's Speech Patterns

Children's speech responses were analyzed acoustically in accordance with the canonical forms that children targeted. That is, all the tokens analyzed for the /s/ category were the responses of children imitating the /s/-initial words during the task and the same for the /ʃ/ category. This method does not depend on whether adults would perceive their productions as more /s/-like or more /ʃ/-like and was used to avoid potential unreliability and bias resulting from adult auditory transcription. Figures 3–5 compare the category emergence patterns between the two language groups for M1, M2, and onset F2 frequency respectively. For each acoustic parameter, the measure was averaged separately over the productions of /s/ and /ʃ/ targets, resulting in each child being represented by two data points (one for /s/ and one for /ʃ/) in each of the figures. These averaged values were plotted against children's chronological age (in months). For each target fricative category (i.e., /s/ or /ʃ/), a simple linear regression line was calculated, with the dependent variable being the mean acoustic values of tokens for the fricative target in each of the acoustic dimensions, M1, M2 or onset F2 frequency, and the independent variable being children's age in months. In addition, a 95% confidence interval was calculated for the regression line of each target in each of the acoustic dimensions to quantify when the regression lines for two fricative targets start to diverge significantly in regression slopes.

The results for the M1 dimension (Figure 3) show that as early as 35 months for English-speaking children, the regression lines for the two target fricatives, /s/ and /ʃ/, begin to show differentiated slopes with no overlap in the 95% confidence interval bands between the two target fricatives. In addition, the two regression lines diverge further in older children. For Japanese-speaking children, such a divergence does not take place until approximately 50 months. Also, the two language groups differ in

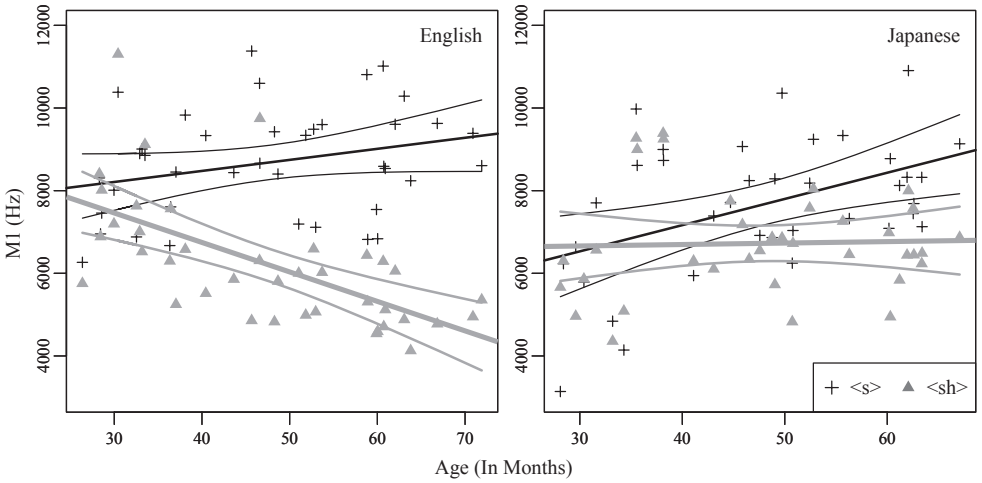

Figure 3. Mean M1 values of all the tokens for each individual child plotted against children's ages for each target fricative.
Note. Best fit lines from linear regression models were overlaid on the data points together with 95% confidence interval bands. For Figures 3–5, the black regression lines and the associated 95% confidence interval bands are for target /s/ and the gray regression lines and 95% interval bands are for target /ʃ/.

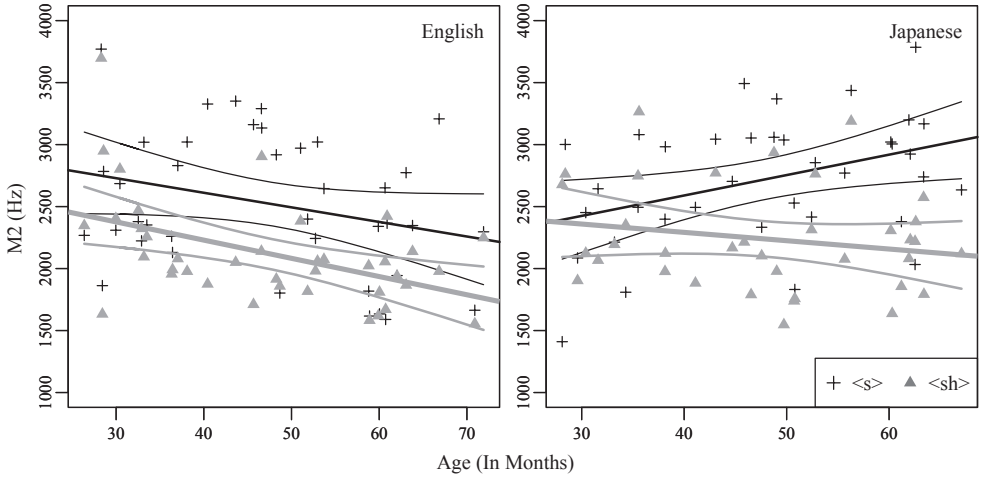

Figure 4. Mean M2 values of all the tokens for each individual child plotted against children's age for each fricative target.
Note. Best fit lines from linear regression models were overlaid on the data points together with 95% confidence interval bands.

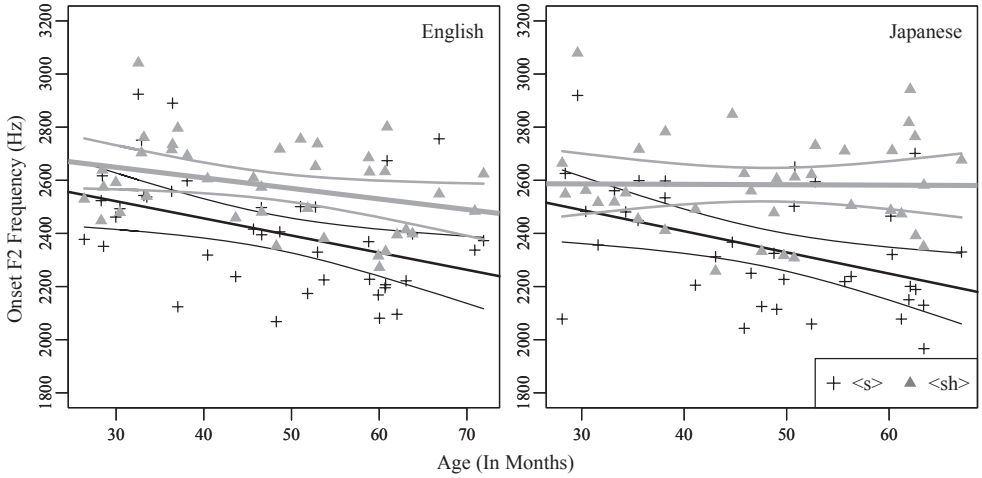

Figure 5. Mean onset F2 frequency values of all the tokens for each individual child plotted against children's age for each fricative target.
Note. Best fit lines from linear regression models were overlaid on the data points together with 95% confidence interval bands.

the magnitude of separation by age 5, with English-speaking children showing a greater degree of separation between the two target fricatives than Japanese-speaking children. A finding of interest involves where the early undifferentiated forms are located in the multidimensional acoustic space. For English-speaking children, the early undifferentiated fricatives center around 8000 Hz: a frequency region intermediate between a mature /s/ and a mature /ʃ/ as produced by 5-year olds, but closer to the /s/ sound. For Japanese-speaking children, the earliest fricative productions center at a lower frequency range around 7000 Hz, occupying an acoustic region closer in value to the /ʃ/ target than to the /s/ target; this is also evident by the steeper regression line for /s/ than that for /ʃ/.

In the M2 dimension, English-speaking children show decreases in M2 values for both targets in older age groups. The slopes for the two targets are nearly parallel, with no clear divergence identified. For Japanese-speaking children, the two target fricatives show a gradual separation. The slope is positive for the /s/ target but negative for the /ʃ/ target. Further, the magnitude of such divergence is comparable to the divergence in the M1 dimension. Again, the early form in the M2 dimension for Japanese children lies intermediate between mature /s/ and mature /ʃ/ productions of 5-year-olds.

The onset F2 dimension for English-speaking children decreases from younger children to older children for both targets. The two regression lines are close to each other and nearly parallel. For

Japanese-speaking children a larger separation is found, beginning at around 40 months. The early undifferentiated form is intermediate between the two target fricatives that a 5-year-old would produce, similar to the pattern in the M2 dimension.

Multiple linear regressions were applied to each acoustic parameter to statistically quantify whether the bestfitting slopes for the two target fricatives were significantly different from each other or not. For each multiple linear regression model, the dependent variables were values of individual acoustic parameters and the independent variables were the two fricative categories (/s/ vs. /ʃ/), children's age in months, and the interaction between the two. A significant interaction between fricative targets and children's age would indicate significant divergence between the two bestfit lines for the two target fricatives. The results were consistent with the observation made in Figures 3–5: Significant interactions between target fricative and children's age were found only in M1 for English-speaking children ($p < .001$ for M1, $p = .288$ for M2, and $p = .262$ for onset F2), but in all three acoustic parameters for Japanese-speaking children ($p < .001$ for M1, $p < .003$ for M2, and $p < .001$ for onset F2). A separate set of multiple linear regressions were applied to words that more than 50% of 2-year-olds are able to produce, according to CDI or JCDI. Similar results were obtained for English-speaking children ($p = .02$ for M1, ps not significant for M2 and onset F2). For Japanese, no /ʃ/-initial words are among the words that more than 50% of 2-year-olds would produce, and therefore a similar analysis could not be carried out.

Discussion

To address the question of how linguistic universality and environmental exposure interact in the course of speech production development, word-initial fricatives produced by English- and Japanese-speaking children were spectrally examined and compared. The differential environmental input was first established by verifying the existence of statistically quantifiable acoustic patterns in adult productions of the shared sounds, /s/ and /ʃ/, for both languages. Despite the common phonological labels assigned to these sounds in both languages, adult speakers of English distinguish /s/ and /ʃ/ mainly in the dimension of M1, that is, where the major energy peaks in the frequency spectrum of the frication noise, while Japanese speakers distinguish /s/ and/ʃ/ in all three acoustic parameters (M1, M2, an onset F2), that is, not only where the mean energy locates but also in the general spectral shape of the noise spectrum as well as the formant values in the transition to the following vowel. Interestingly, children's speech patterns exhibit systematic divergences in the same dimensions that adult speakers use to distinguish the two fricative categories on a language-specific basis. That is, a developmental separation occurred only in M1 for English-speaking children, but in all three acoustic dimensions (M1, M2 and onset F2) for Japanese-speaking children.

In addition, children's speech patterns also resemble those of adults' in the magnitude of importance of each dimension. For M1, a dimension relevant for both English and Japanese, the degree of separation is much greater for English-speaking children than that for Japanese-speaking children. This is in accordance with adult norms where M1 alone contributes to all the differentiation between /s/ and /ʃ/ for English-speaking adults, whereas it only contributes partially to the /s/–/ʃ/ distinction for Japanese-speaking adults. This parallel is not surprising given that statistical regularities in speech can be detected by infants through exposure to ambient language (Maye, Werker, & Gerken, 2002; Saffran, Aslin, & Newport, 1996). The current research demonstrates how children are attuned to language-specific relevant acoustic dimensions in mastering speech production.

Finally, cross-language differences were found where the early undifferentiated forms are located in the acoustic space. For the M1 dimension, where both language groups show gradual categorical separation, the early undifferentiated form occupies a frequency region around 8000 Hz for English-speaking children, and this is approximately 1000 Hz higher than that for Japanese-speaking children. The higher frequency range in English-speaking children's early fricative productions is likely to yield a percept of /s/ if adults were forced to transcribe them. Similarly, the lower frequency range for the early form produced by Japanese-speaking children is more likely to give rise to the /ʃ/ percept. This is consistent with reports from transcription studies of earlier acquisition of /s/ in English but earlier acquisition of /ʃ/ in Japanese (Nakanishi et al., 1972; Smit et al., 1990). The differences in the early forms corroborate with the results of a recent study on English- and Japanese-speaking children's fricative development that the emergence of the same two categories among individual children's speech are below the perceptual threshold of adults (i.e., covert contrast; Li et al., 2009). In that study, some 2- to 3-

year-olds were found to produce statistically quantifiable differences between /s/ and /ʃ/ that were identified as homophonous by adult transcribers. It is important to note that such subtle differences emerge in the acoustic range corresponding to the adult /s/category for English-speaking children and to the /ʃ/ category for Japanese-speaking children. It is also worth noting that the results from the Li et al. (2009) study suggests that perceptual judgments underestimate children's competence, which is consistent with the current study: English-speaking children showed statistically quantifiable separation as early as 35 months, while large-scale transcription studies report an age of acquisition around 48 months (Smit et al., 1990; Templin, 1957).

The tendency to produce more /s/-like forms in English-speaking children and more /ʃ/-like forms in Japanese-speaking children may be attributed to how frequently children encounter each sound in their native linguistic environment. For example, In English, /s/ is about 6 times more frequent than /ʃ/ (Edwards & Beckman, 2008). A recent study on English-speaking adult fricative perception (Li, Munson, Edwards, Yoneyama, & Hall, 2011) revealed a bias to categorize /s/ when judging children's speech; a similar bias toward /s/ has also been shown in children's perception (Nittrouer & Miller, 1997b). Such uneven distribution in sound frequency could affect children's phonological representation, which could impact their speech productions in two ways. First, it allows more auditory exposure to the sound /s/ than to the sound /ʃ/, or even creates a quasi-unimodal distribution with a mean toward the/s/ sound instead of a bimodal distribution. Second, more experience with words beginning with /s/ could allow children to extract away the sound segment earlier than other, less commonly used speech sounds. In Japanese, although /s/ and /ʃ/ are about equally frequent, caregivers frequently palatalize their /s/ sound as a sound symbolism for smallness or cuteness in addressing young children, a process resulting in more /ʃ/-like input than suggested by lexicon (Chew, 1969). Palatalization in child-directed speech in Japanese could also account for the apparent discrepancy between fewer number of familiar /ʃ/-initial words in JCDI and the earlier emergence of /ʃ/-like forms in children's speech. Such results thus indicate the necessity and importance of describing linguistic input to children below the phonological level.

The current study provides support for the continuous view of speech production development, echoing previous studies on child speech that used instrumental techniques (Boysson-Barties et al., 1989; Kuhl & Meltzoff, 1996; Macken & Barton, 1980; Nittrouer, 1995). An important question then is: If acquiring speech sounds is primarily a process of differentiation, then what prevents children from producing distinctive categories at the early phase and what drives the subsequent separation? The maturation of children's anatomical structures and motor control of vocal organs could be primarily responsible for such a process. First, young children's vocal tracts are shorter, broader, and flatter in comparison to those of adults, and they do not have enough space to manipulate the tongue freely (Kent, 1992; Vorperian et al., 2009). Second, in order to achieve the fine positioning necessary for fricative articulations, children also need to gain separate control of different parts of the tongue, which requires years of practice. The inability to independently manipulate the tongue apex and blade has been found in some children with articulation disorders (i.e., ''undifferentiated lingual gesture''; Gibbon, 1999). Similar insufficient motor control may cause the early undifferentiated forms found in children's speech reported in the current study. One possible avenue for future research is to observe changes in motor control across development by directly measuring tongue movements using techniques such as ultrasound.

The lack of differentiation in early fricative productions may also be due to perceptual confusion or due to incomplete formation of perceptual categories. Although infants could discriminate /s/ from /ʃ/ as early as 4 months of age (Eilers & Minifie, 1975), children at age 2 still had difficulty in applying this knowledge in a word-learning task (Barton, 1980). Further, the fine attunement to the defining acoustic dimensions and their associated weights in fricative perception is not perfected until school years. In a series of fricative perception experiments testing children aged 3, 5, and 7, Nittrouer and colleagues (Nittrouer, 1992; Nittrouer & Miller, 1997a) created a continuum of fricative segments with /s/ and /ʃ/ being the end points by systematically incrementing the mean frequency of the fricative noise spectrum in individual steps. These synthetic fricative segments were then combined with natural vowels with transitions appropriate for /s/ or for /ʃ/. When asked to identify whether the resulting word is *Sue* or *shoe*, younger children relied more on the transitional information, whereas older children relied more on the frication mean frequency. It is not until age 7 that children approximate the weighting scheme that adult speakers use. Such progression from attending to less relevant acoustic dimensions to the defining dimensions in perception may

underlie the gradual differentiation in fricative productions demonstrated in this article.

One limitation of the current study is that there was incomplete control for word familiarity in the word stimuli used to elicit children's sound productions. Word familiarity entails both familiarity with the form and familiarity with the meaning (Cordier & Le Ny, 2005). It is out of the scope of this article to examine the extent to which different degrees of word familiarity affect the fine phonetic details in each child's articulation. However, the design of the current experiment does minimize the influences of word familiarities by carefully selecting the pictures and normalizing the auditory prompts by asking naïve native speakers to repeat the prompts. Further, the test sounds were always at the most salient position (i.e., word-initial position) to enhance the likelihood of correct imitation to counteract the smaller short-term memory capacity in children. The same procedure has actually been used to elicit nonsense words from young children and has been demonstrated to be successful (Munson, Edwards, & Beckman, 2005). The comparison of articulation performance between real words and nonsense words can address the question of whether word familiarity will affect children's speech production at all. This analysis is now under way.

To conclude, differentiation processes for the shared sounds /s/ and /ʃ/ are common to both English-speaking and Japanese-speaking children during articulation development. What differs between English-speaking and Japanese-speaking children is where and how the differentiation takes place in the multidimensional acoustic space. These differences can be readily attributed to the specific language that each child was exposed to. Therefore, it is clear that long before children are able to produce sound distinctions between /s/ and /ʃ/ in an adult-like manner, they already fine tune their motor control towards the mother tongue.

References

Akamatsu, T. (1997). *Japanese phonetics: Theory and practice*. Newcastle, UK: Lincom Europa.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Barton, D. (1980). Phonemic perception in children. In J. F. K. G. H. Yeni-Komshian & C. A. Ferguson (Eds.), *Child phonology: Vol. 2. Perception* (pp. 97–114). New York: Academic Press.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version praat 4.3.07). Retrieved December 1, 2005, from http://www.praat.org/

Boysson-Barties, B. de., Halle, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, *16*, 1–17.

Chew, J. J. (1969). The structure of Japanese baby talk. *Journal-Newsletter of the Association of Teachers of Japanese*, *6*, 4–17.

Cordier, F., & Le Ny, J. (2005). Evidence for several components of word familiarity. *Behavior Research Methods*, *37*, 528–537.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying phonological acquisition. *Clinical Linguistics and Phonetics*, *22*, 939–958.

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research*, *18*, 158–167.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague, Netherlands: Mouton.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, *21*, 95–115.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, *84*, 115–124.

Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological factors or sociophonetic ones. In S. Fuchs, M. Toda, & M. Zygis (Eds.), *An interdisciplinary guide to turbulent sounds* (pp. 281–302). Berlin: Mouton de Gruyter.

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

Gibbon, F. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, *42*, 382–397.

Halle, M., & Stevens, K. N. (1997). The postalveolar fricatives of Polish. In S. Kiritani, H. Hirose, & H. Fujisaka (Eds.), *Speech production and language: In honor of Osamu Fujimura* (Vol. 13, pp. 176–191). New York: Mouton de Gruyter.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Jakobson, R. (1960). *Child Language, aphasia, and phonological universal*. The Hague, Netherlands: Mouton(Original work published 1941).

Jesus, L. M. T., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, *30*, 437–464.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*, 1252–1263.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, *19*, 421–447.

Kent, R. (1992). *The biology of phonological development*. Timonium, MD: York Press.

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, *5*, 7–23.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*, 2425–2438.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK: Blackwell.

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

Li, F., Munson, B., Edwards, J., Yoneyama, K., & Hall, K. (2011). Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. *Journal of the Acoustical Society of America*, *129*, 999–1011.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in english: A study of the voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*, 41–74.

MacNeilage, P. F., & Davis, B. L. (1990). Acquisition of speech production: Achievement of segmental independence. In W. I. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 55–68). Dordrecht: Netherlands.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research*, *48*, 61–78.

Nakanishi, Y., Owada, K., & Fujita, N. (1972). Koon kensa to sono kekka no kosatsu [Results and interpretation of articulation tests for children]. *RIEEC Report* [Annual Report of Research Institute Education of Exceptional Children, Tokyo Gakugei University], *1*, 1–41.

Narayanan, S. S., Alwan, A. A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America*, *98*, 1325–1347.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, *20*, 351–382.

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research*, *36*, 959–972.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America*, *97*, 520–530.

Nittrouer, S., & Miller, M. E. (1997a). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America*, *102*, 572–580.

Nittrouer, S., & Miller, M. E. (1997b). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, *101*, 2253–2266.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Scobbie, J. M. (1998). Interactions between the acquisition of phonetics and phonology. In D. H. M. C. Gruber, K. Olson, & T. Wysocki (Ed.), *The 34th annual regional meeting of the Chicago Linguistic Society* (Vol. 2, pp. 343–358). Chicago: Chicago Linguistics Society.

Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics*, *19*, 409–424.

Smit, A. B., Hand, L., Frieilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*, 29–36.

Stoel-Gammon, C., Williams, K., & Buder, E. (1994). Cross-language differences in phonological acquisition: Swedish and American /t/. *Phonetica*, *51*, 146–158.

Tabain, M., & Watson, C. (1996). Classification of fricatives. In P. McCormack & A. Russell (Eds.), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology* (pp. 623–628). Adelaide, Australia: Australian Speech Science and Technology Association.

Templin, M. (1957). *Certain language skills in children*. Minneapolis: University of Minnesota. Minnesota.

Toda, M. (2007). Speaker normalization of fricative noise: Considerations on language-specific contrast. *Proceedings of the XVI International Congress of Phonetic Sciences* (pp. 825–828). Dudweiler, Germany: Pirrot.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., et al. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of the Acoustical Society of America*, *125*, 1666–1678.

Yasuda, A. (1970). Articulatory skills in three-year-old children. *Studia Phonologica*, *5*, 52–71.

Appendix

English stimuli:

Target fricative	Vocalic context	Words	Phonetic transcription	Proportion of 2-year-olds producing (CDI)
/s/	A	sun	/sʌn/	58.9
		sauce	/sas/	34.6
		soccer	/sakɚ/	
	E	same	/sem/	24.3
		safe	/sef/	
		seven	/sɛvn/	
	I	seashore	/siʃɔ/	
		seal	/sil/	
		sister	/sɪstɚ/	36.4
	O	soldier	/soldʒɚ/	
		soak	/sok/	
		sodas	/sodaz/	53.3
	U	super	/supɚ/	
		suitcase	/sutkes/	
		soup	/sup/	58.9
/ʃ/	A	shovel	/ʃʌvl/	43
		shark	/ʃak/	
		shop	/ʃap/	(shopping) 51.4
	E	shell	/ʃel/	
		shepherd	/ʃɛpɚd/	
		shape	/ʃep/	
	I	sheep	/ʃip/	52.3
		ship	/ʃɪp/	
		shield	/ʃild/	
	O	shore	/ʃɔ/	
		show	/ʃo/	25.2
		shoulder	/ʃodɚ/	36.4
	U	chute	/ʃut/	
		sugar	/ʃugɚ/	
		shoe	/ʃu/	90.4

Japanese stimuli:

Target fricative	Vocalic context	Phonetic transcription	Gloss	Proportion of 2-year-olds producing (JCDI)
/s/	A	/sakana/	''fish''	67.9
		/sakura/	''cherry blossom''	
		/saru/	''monkey''	60.4
	E	/semi/	''cicada''	
		/senaka/	''back''	27.3
		/sensɛ:/	''teacher''	32.1
	O	/sok:usu/	''socks''	7.0
		/so:se:dʒi/	''sausage''	
		/sora/	''sky''	21.7
	U	/sudzume/	''sparrow''	
		/suika/	''watermelon''	45.5
		/suna/	''sand''	22.3