The Development of Voiceless Sibilant Fricatives in Putonghua-speaking Children

Fangfang Li

Department of Psychology, University of Lethbridge, Alberta, Canada

Benjamin Munson

Department of Speech-Language-Hearing Sciences, University of Minnesota, Minnesota, U.S.A.

Abstract

Purpose: The aim of present study is 1) to quantify the developmental sequence of fricative mastery in Putonghua-speaking children, and discuss the observed pattern in relation to existing theoretical positions, and 2) to describe the acquisition of the fine-articulatory/acoustic details of fricatives in the multi-dimensional acoustic space.

Method: Twenty adults and ninety-seven children participated in a speech production experiment, repeating a list of fricative-initial words. Two independent measures were applied to quantify the relative sequence of fricative acquisition: auditory-based phonetic transcription and acoustics-based statistical modeling. Two acoustic parameters, fricative centroid frequency and F2 onset, were employed to index tongue body and tongue tip development respectively.

Results: Both transcription and statistical modeling of acoustics yielded the sequence of /ɕ/ → /ʂ/ → /s/. Acoustic analysis further reveals gradual separation in both acoustic dimensions, with the initial undifferentiated form ambiguous between /ɕ/ and /ʂ/.

Conclusions: The observed sound acquisition order was interpreted as reflecting a combined influence of both oromotor maturation and language-specific phoneme frequency in Putonghua. Acoustic results suggest a maturational advantage of the tongue body over the tongue tip during fricative development.

Children learn to produce speech sounds at varying rates. The relative sound acquisition sequence reflects the combined forces of common oromotor and perceptual maturation that constrain the developmental processes of all children, and important environmental influences exerted from the ambient language (de Boysson-Bardies, Halle, Sagart, & Durand, 1989; Edwards & Beckman, 2008; Ingram, 1988a; Li, 2012; Pye, Ingram, & List, 1987; Ferguson & Farwell, 1975; Jakobson, 1941/1968; Kent, 1992; Locke, 1983). In this study, we describe the acquisition pattern of voiceless sibilant fricatives in Putonghua-speaking children using both native-speaker phonetic transcription and statistical modeling based on acoustic analysis. The chronology we identify of the emergence of these sounds serves to delineate the relative contribution of three factors proposed by previous theoretical frameworks: cross-language phoneme frequency, language-specific phoneme frequency, and articulatory maturation.

Prior theoretical work on factors influencing phonological development

Cross-language phoneme frequency is the legacy of the pioneering child phonologist, Roman Jakobson. In his seminal monograph *Child Language, Aphasia, and Phonological Universals,* Jakobson (1941/1968) proposed that the distribution of phoneme types among the world languages should predict the sound acquisition sequence. For example, the frequently observed acquisition advantage of vowels, stops, and nasals over fricatives and affricates was attributed to their common occurrence across languages. Jakobson (1941/1968) proposed the "fronting universal" to account for the development of consonant place of articulation. This hypothesis, later elaborated by Locke (1983), claims an early acquisition of the anterior places of articulation as compared to the posterior ones, drawing evidence from [t]-for-/k/ or [s]-for-/ʃ/ substitution patterns produced by children from diverse language backgrounds.

Such claims of universal tendencies were later challenged by more recent cross-language studies of phonological acquisition, suggesting the importance of phonemes' language-specific frequency of occurrence (Ingram, 1988a; Pye, Ingram, & List, 1987). In an investigation of phonological development of five children speaking Quiché, the usually late-acquired /tʃ/ was found to be among the earliest emerged sounds (Pye, Ingram, & List, 1987). This is true despite the fact that /tʃ/ has a low cross-language frequency, occurring in only 4.21% of the 451 languages in the UCLA Phonological Segment Inventory Database (UPSID, Maddieson, 1984; Ladefoged & Maddieson, 1996). Pye and colleagues attributed this acquisition pattern to the high frequency of occurrence of /tʃ/ relative to other phonemes across the Quiché lexicon. This in turn suggests that the lexical statistics in the developing child's language environment can affect acquisition substantially. One caveat to using UPSID for references on phoneme frequencies is that they are subject to the auditory judgments of field researchers and their language-specific phonemic inventory. The limitation of the transcription-based study will be discussed in greater detail in the section "Methodological challenges".

It is seemingly axiomatic that anatomical maturation and the development of oromotor control contribute to the process of sound acquisition, although the exact mechanism through which anatomical and oromotor factors shape the sequence of acquisition remains elusive. One principle of the development of gross and fine motor control, the *proximal-distal principle*, has been invoked to explain some facts about speech-sound development. This principle explains the developmental sequence of control that progresses from large-muscle (i.e., arms and legs) to small-muscle (i.e., hands and toes) use, and from central (i.e., palm) to peripheral (i.e., fingers) body parts (Butterworth, Verweij, & Hopkins, 1997; Irwin, 1933; McBryde & Ziviani, 1990; Wallace & Whishaw, 2003). A few researchers have speculated that the same principle could apply to tongue muscle maturation, with children gaining earlier control over the tongue body than the components of the tongue such as the tongue tip (Li, 2008; Gibbon, 1999). Such a hypothesis could readily explain a peculiar acquisition phenomenon concerning the alveolopalatal fricative /ɕ/. This sound is often reported to emerge earlier than other voiceless sibilant fricatives in children's speech in many languages (Nakanishi et al., 1972 for Japanese; Zharkova, 2005 for Russian, and Ingram, 1988b, for Polish), despite its rarity both across languages and within some languages. The main difference between /ɕ/ and other sibilant fricatives lies in its utilization of the large, more central muscles controlling the dorsum as the major articulator as compared to the use of the smaller, more peripheral muscles that control the tongue tip in /s/ or /ʃ/.

Methodological challenges

All above-mentioned theoretical controversies on the roles played by various factors rely on the observed sound acquisition sequence. Thus, reliable descriptions of children's speech production patterns are fundamental to resolving controversies about phonological development. However, the reliability of the classical auditory-based transcription method has been increasingly challenged due to its coarse-grained nature, and to the potential bias introduced through perceptual judgment (Gibbon, 1999; Kent, 1992; Ladd, 2011; Munson, Johnson, & Edwards, 2012; Munson, Edwards, Schellinger, Beckman, & Meyer, 2010; Scobbie, Gibbon, Hardcastle, & Fletcher, 2000). People perceive speech in a categorical fashion, experiencing perceptual equivalence for variation within a phoneme's boundary ("categorical perception", Liberman, Harris, Kinney, & Lane, 1961). When children produce two sounds in a range corresponding to a single adult sound category, transcribers tend to overlook the subtle distinction, a phenomenon termed "covert contrast" which was supported by a growing body of literature that uses instrumental analysis (Baum & McNutt, 1990; Li, Edwards, & Beckman, 2009; Mackon & Barton, 1980; Scobbie, Gibbon, Hardcastle, & Fletcher, 2000).

Another issue with the search for sound acquisition order lies in the assumption that speech sound development unfolds in a discrete fashion, as explicitly dictated by Jakobson's "implicational universals" or "laws of irreversible solidarities". Importantly, this assumption has not been upheld by instrumental studies. The developmental process of speech sounds revealed by instrumental analysis is more gradual than what transcription studies have suggested (Nittrouer, 1995; Li, 2012; Mackon & Barton, 1980; Scobbie, Gibbon, Hardcastle, & Fletcher, 2000).

On the other hand, although instrumental analysis offers rich information on children's speech profile, it has its own limitations. The validity of the acoustic measurements is dependent on researchers using the parameters that best reflect the child's articulation, as well as the auditory parameters that most strongly predict listeners' perceptual judgments. Furthermore, the results of instrumental analysis still require human interpretation. These and other factors explain why few studies report the acoustics of children's speech on a scale comparable to what most normative studies report with phonetic transcription.

Purposes

The purpose of the present study is two-fold. The first is to solidify the chronology of sound emergence pattern through the combination of the two widely used analytical tools (i.e., transcription and acoustic methods) in a relatively large sample of Putonghua-speaking children. The employment of the acoustic analysis allows for a more in-depth examination of children's acquisition of the fine-grained phonetic/articulatory details that are generally thought to be below listeners' perceptual thresholds. And the transcription method provides a frame for the interpretation of the acoustic patterns. Secondly, we can use the data on the acquisition of voiceless fricatives in Putonghua to examine the relative contribution of the factors proposed by various theoretical frameworks. Putonghua, standard Mandarin spoken in mainland China, is particularly suitable for investigating fricative acquisition as it contains a rich inventory of three voiceless sibilant fricatives: alveolar/dental /s/, alveolopalatal /ɕ/, and retroflex /ʂ/ (Ladefoged & Wu, 1984; Lin & Wang, 1992). Note that the retroflex fricative in Mandarin is different from those in other languages in that it does not involve the curling back of the tongue tip, and therefore termed as "flat retroflex" (Ladefoged & Wu, 1984; Ladefoged & Maddieson, 1996). The following section reviews the literature on Putonghua fricatives and makes predictions about the Mandarin-speaking children's fricative development in relation to each of the three factors mentioned previously: cross-language phoneme frequency, phoneme frequency across the Putonghua lexicon, and articulatory maturation.

Cross-language phoneme frequencies. A calculation of phoneme type frequencies using the UPSID database reveals the prevalence of /s/ occurring in 197 languages in contrast to the rarity of /ɕ/ (21 languages) and /ʂ/ (23 languages). This predicts that /s/ should be acquired prior to the two posterior fricatives. Jakobson's implicational universals, in particular the fronting universal, similarly predict /s/ to emerge first in Putonghua-speaking children.

Phoneme frequencies in Putonghua. Tsoi (2005) calculated the Mandarin phoneme frequencies on the basis of phonetic transcriptions of the Lancaster Corpus of Mandarin Chinese (LCMC). This yielded the highest frequency for /ʂ/ (32357 times), followed by /ɕ/ (23199 times) and lastly /s/ (6641 times). LCMC is a Chinese corpus sampling 15 written text categories such as news, literary texts, academic prose, and official documents, published in mainland China around the year of 1991 for a total of approximately one million words (McEnery & Xiao, 2004).

Similar findings were obtained by Li (2008), which reported the log frequency of voiceless sibilant fricatives in Putonghua using the CALLHOME Mandarin Chinese lexicon retrieved from the online Linguistic Data Consortium (LDC) catalog (Huang, Bian, Wu, and McLemore, 1997). The corpus consists of 44,405 words of conversational speech. Li (2008) calculated the log phoneme frequency by taking the log ratio of the number of words beginning with a fricative to the total number of words in the corpus. The higher log frequency is indicative of high phonemic frequency, and reported log frequencies are highest for /ʂ/ (-2.7), followed by /ɕ/ (-2.8) and /s/ (-3.9).

Based on these two studies, if language-specific phoneme frequency plays a decisive role in determining the acquisition order of fricatives, /ʂ/ and then /ɕ/ will be mastered earlier than /s/.

Articulation characteristics of Putonghua fricatives. Both /s/ and /ʂ/ are articulated with the tongue apex (Hu, 2008; Ladefoged & Wu, 1984; Lee, 1999). The distinction between /s/ and /ʂ/ primarily lies in the relative constriction point in the oral cavity – the narrowest constriction for /s/ is around the alveolar ridge or right behind the upper incisors while that for /ʂ/ is posterior to the alveolar ridge (Ladefoged & Wu, 1984; Lee, 1999; Hu, 2008). In contrast, it is the tongue dorsum instead of the apex that is employed for the articulation of /ɕ/. Specifically, the tongue predorsum is elevated towards the hard palate. This creates a long palatal passage to channel forced air (Ladefoged & Wu, 1984; Lin & Wang, 1992; Toda & Honda, 2003). If tongue maturation follows the sequence of dorsum then apex, as the proximal-distal principle predicts, then control over the distinct muscles recruited for making /ɕ/ should be acquired prior to the control for those that are used to produce /s/ and /ʂ/. If tongue maturation is the only/primary factor in determining Mandarin fricative acquisition order, then /ɕ/ is expected to precede the sounds /s/ and /ʂ/.

Previous research on Putonghua fricative acquisition. Few previous studies have documented Putonghua-speaking children's sound development. This prevents a clear evaluation of the relative contribution of the three factors. Most existing work (i.e., Jeng, 1979; Li & Tompson, 1977; Lin & Peng, 2003) was conducted in Taiwan. In Taiwan, children acquire a different variety of Mandarin, Taiwan Guoyu, which is highly influenced by another language spoken in Taiwan, Taiwanese. Taiwan Guoyu tends to merge /s/ with /ʂ/, as Taiwanese does not have the /ʂ/ sound (Shih & Kong, 2011).

The handful of studies examining the development of Putonghua in children yield mixed results regarding the developmental sequence of fricatives (i.e., Li, Zhu, & Dodd, 2002; Si, 2005; Xu, Yang, & Qi, 2010). In a longitudinal study investigating the speech development of four Putonghua-speaking children, Li, Zhu, and Dodd (2002) found two children produced /ɕ/ first, while the other two children produced /s/ first. Zhu and Dodd (2000) conducted a large-scale cross-sectional study on 129 children aged 1 to 4 years. The study revealed an earlier mastery of /ɕ/ (2;7~3;0) followed by /s/ (4;1~ 4;6), with /ʂ/ >4;6) being acquired last. However, since Zhu and Dodd's study targeted the entire Putonghua consonant inventory, the sampling of the three voiceless sibilant fricatives was both sparse and unbalanced: /ɕ/ and /ʂ/ were elicited 8 and 4 times respectively in word-initial position, while /s/ was only elicited once in word-medial position. More systematic study is called for to determine the developmental pattern of Putonghua fricatives.

Acoustic measurement. Phonetic transcription has been the main investigation tool for previous research on Mandarin children's phonological development. The present study combines both transcription and acoustic analysis to provide two independent measurements to quantify fricative learning robustly. The acoustic analysis can reveal fine-grained articulation details that would fall within even a very skilled phonetic transcriber's categorical boundaries. Separate acoustic parameters can also index different aspects of articulation to inform motor control development.

Two acoustic parameters proven effective in previous research on Mandarin Chinese will be applied in the acoustic analysis of the present study: 1) centroid frequency calculated from the middle portion of fricative noise and 2) Second formant (F2) frequency taken at the onset of the following vowel. Centroid frequency is the mean frequency of fricative noise spectrum (Forrest, Weismer, Milenkovic, & Dougall, 1988), which increases as the constriction point of a fricative moves towards a more anterior position in the oral cavity (Jongman, Wayland, & Wong, 2000; McGowan & Nittrouer, 1988; Shadle & Mair, 1996). Consequently, the centroid frequency of Mandarin fricatives is expected to vary in the order of /s/ > /ɕ/ > /ʂ/ (Svantesson, 1986; Li, 2008). The centroid measure can index the location of the lingual constriction forward or backward in the mid-sagittal plane of oral cavity.

The second measure, F2 onset frequency, is inversely correlated with the length of the back cavity, and /ɕ/ is expected to exhibit the highest value due to its distinct tongue posture (Halle & Stevens, 1997; Li, 2008; Stevens, Li, Lee, & Keyser, 2004). Unlike /s/ or /ʂ/ whose tongue shape is relatively flat and whose back-cavity length is determined by where the tongue tip is located, the whole tongue dorsum is bunched up and raised towards the hard palate in producing /ɕ/. As a result of this posture, the back cavity of /ɕ/ is much reduced in comparison to that of /s/ or /ʂ/. Between /s/ and /ʂ/, the length of the back cavity is slightly longer for /ʂ/, owing to its retracted tongue tip position, but the length of the back cavity is longer for both sounds than for /ɕ/. F2 onset thus offers a way to index the upward or downward movement of the tongue body in the coronal plane of the oral cavity.

Methods

Participants

Ninety-seven children aged 2 to 5 years were tested in Songyuan, Jilin province, China. No child spoke languages other than Putonghua or other Chinese dialects, and their parents were native speakers of Putonghua. All children tested had normal hearing and passed a hearing screening test using otoacoustic emissions at 2000, 3000, 4000, and 5000 Hz. No child tested had any reported speech, language or hearing problems, according to parents' or teachers' report. In addition, 20 adults from the same region (gender-balanced, aged 18-30 years) were tested to serve as the baseline for comparison with children's speech productions. Table 1 displays the breakdown of participants as a function of their gender and age.

Insert Table 1 about here

Task & Materials

Children were tested individually in a quiet room of a daycare center. They were seated in front of an IBM laptop computer, with an AKG microphone placed approximately 20 cm from their mouth. Pictures representing stimulus words were displayed in the middle of the computer screen and were presented simultaneously with audio prompts. The whole procedure was facilitated through the "Show and Play" program (Edwards & Beckman, 2008), which adds an image of a duck climbing a ladder at the left margin of the screen. Children were told to play a computer game by listening to the word the computer says first and then repeat it back to the microphone. They were also told that whenever they repeated a word, the duck would climb up one step, and they would win the game if they help the duck to climb to the top of the ladder. A practice session was offered prior to the testing to familiarize participants to the task. Each session lasts approximately 5 to 10 minutes. For 2-year-olds, a break was provided in the middle of the session, and/or stickers were used to prompt them for a response. The procedure for testing adults was similar to that of children, except that they were informed of the purpose of the experiment as being primarily targeted at child speakers.

The test materials were words beginning with fricative-vowel sequences. The words were selected on the basis of picturability and familiarity to children. Word familiarity was ensured by asking parents to check the familiar words off the list of words used in the experiment. Not all the stimulus words are familiar to 2-year-olds. However, no significant difference was found between familiar and unfamiliar words with respect to the transcribed accuracy, and therefore the two groups of words were collapsed for later analysis.

Each fricative was sampled in word-initial positions (as Mandarin has no coda fricatives) in 16 words and was followed by one of the five vowels, /a/, /i/, /u/, /ɛ/, and /o/ (see Appendix A for the entire word list). Due to phonotactic constraints, no words were included for the sequence */sɛ/, */ʂɛ/, and */ɕu/. Thus, a total of 4656 tokens (16 × 3 targets × 97 children) were elicited from children. However, due to circumstances such as productions with unintended words or unrecognizable words, 31 tokens were removed from the transcription analysis, yielding a total of 4625 tokens (1540 for /s/, /, 1542 for /ʂ/, and 1543 for /ɕ/). Furthermore, with respect to acoustic analysis, tokens with deletion and manner errors (stopping and affrication) as judged by the transcriber or those overlapping with background noise were excluded from the acoustic analysis, resulting in 4240 remaining tokens for children. For adults, 960 tokens (16 × 3 targets × 20 adults) were elicited and included in the acoustic analysis.

Procedure

Speakers' productions were digitally recorded to a Marantz PMD 660 portable recorder with a 44100 Hz sampling rate and 16-bit digitization. These audio recordings were subsequently submitted to transcription and acoustic analyses. Praat (Boersma, & Weenink, 2005) was used for raw data processing and transcription. As mentioned previously, a total of 4625 tokens were transcribed and included in the statistical analysis. The first author, a native speaker of Putonghua and trained phonetician who is also from the same region of China as the participants, transcribed all initial target fricatives using 1 (for correct productions) and 0 (incorrect). For mispronounced tokens, the transcriber also noted the error patterns. A second native speaker of Putonghua, who was phonetically trained, independently transcribed 20% of the data (4 two-year olds, 5 three-year olds, 5 four-year olds, and 5 five-year olds), and the phoneme-by-phoneme inter-rater reliability was 85% for /ɕ/, 89% for /ʂ/, and 98% for /s/, with a mean of 92%.

For acoustic analysis, F2 onset was measured at or after the end of the fricative noise, which was defined as first zero crossing in the upswing voicing cycle of the following vowel. Centroid frequency was calculated from a Multitaper spectrum based on the middle 40-millisecond slice of each fricative segment using the Multitaper package (Rahim, 2010) in R (R Development Core Team, 2011). Each Multitaper spectrum was high-pass filtered (above 1000 Hz) to eliminate potential low-frequency noise such as wind blowing or door opening/closing.

Results

Transcription analysis

Table 2 presents the number and proportion of tokens that were judged to be correct, separated by age group and vowel contexts. Although the actual accuracy rates (i.e., weighted proportions) vary slightly from vowel to vowel, the overall pattern was clear: for 2- and 3- year-olds, the sound /ɕ/ was slightly more accurate than /ʂ/. Both /ɕ/ and /ʂ/ show much higher accuracy rates than /s/. Around age 4, the gap between /s/ and the other sound starts to close, and by age 5, all three sounds have rates higher than 0.9.

A repeated measures ANOVA was constructed to test the effect of age, fricative consonants and vowels on the accuracy rates. The dependent variable was the rationalized arcsine transformed values of the transcribed accuracy for each fricative token. The rationalized arcsine transformation alleviates the floor and ceiling effects commonly present in proportional data (Studebaker, 1985). The independent variables were fricative category (within-subject; three levels: /s/, /ɕ/, and /ʂ/), children's age (between-subject; four levels: 2, 3, 4, and 5), vowel contexts (within-subject; five levels: /a/, /i/, /u/, /ɛ/, /o/), as well as the interaction terms between the three independent variables. A main effect of fricative category ($F(2,184) = 38.99$, $p<0.001$, partial $\eta^2=0.14$) as well as a main effect of age ($F(3, 91) = 66.54$, $p<0.001$, partial $\eta^2=0.40$) were found. There was a significant interaction between age and fricative category ($F(6,184) = 4.33$, $p<0.001$, partial $\eta^2=0.04$), which indicates that specific fricative productions differ for age. This interaction is illustrated by comparing the points (for means) and the standard deviation bars in Figure 1: in general, children improve their articulation accuracy for all three fricatives over the age range we studied, but the pace of this improvement differs for the fricative categories. While /ɕ/ and then /ʂ/ are more accurate than /s/ at age 2 and 3, such an advantage is lost at age 4 and 5 when /s/ is produced comparably accurately.

No significant overall effect of vowel was found ($F(4, 372) = 2.11$, $p = 0.08$). No interaction was found between fricative and vowel ($p=0.18$), nor was there an interaction among fricative, vowel, and age ($p=0.48$), suggesting the minimal role of vowel context played in the acquisition of these fricatives.

Insert Table 2 about here

Insert Figure 1 about here

As shown in Table 2, the overall mean accuracy rates were 0.60 for /s/ (926 correct tokens out of 1540), 0.76 for /ʂ/ (1171 correct tokens out of 1542), and 0.88 for /ɕ/ (1362 correct tokens out of 1543). In total, 3459 tokens out of 4625 were judged to be correct. The remaining 1166 tokens were transcribed as incorrect productions and the error patterns were noted. Table 3 shows the distribution of error types for each of the three fricatives (excluding the 6 or 7% of errors that were not attested in at least 1% of the errors for that fricative). Errors could be deletions (between 3 and 6% of the errors for each fricative type) or substitutions of some other consonant, such as a different fricative, some kind of stop, or an affricate. Of great interest to note is that the majority of errors are substitutions using other fricatives, in particular, the other voiceless sibilants. For example, /s/ was primarily mispronounced as /ʂ/ or /ɕ/. Similarly, /ɕ/ was frequently substituted by /s/ or /ʂ/. In addition, despite the early emergence of /ɕ/, it is not the primary sound that children used to substitute /s/ or /ʂ/, whereas mispronunciations for /s/ and /ʂ/ are primary substitutions for one another. This suggests that the muscles used for /ɕ/ versus /s, ʂ/, or the motor control program for /ɕ/ versus /s, ʂ/, are likely to be mutually incompatible.

Another fact to note in Table 3 is the percentage agreement between the first and the second native-speaker transcriber for each of the transcribed sound category. Despite the high overall inter-rater reliability (92%), variations exists for the percentage of agreement for different transcribed categories, with the lowest being the /ɕ/-to-[tɕʰ] error (57%). The degree of disagreement revealed by the table suggests the ambiguous nature of children's speech production, which further points to the necessity of complementing the transcription analysis with acoustic analysis of children's speech.

Insert Table 3 about here

Table 4 lists the age breakdown of the distribution of the top three error types of each fricative target. The table illustrates that the percentage of errors declines as children age, except for the case of /ʂ/-to-[s] substitution where the highest percentage of errors were found in age 3 and the errors distribute equally across age groups. Such a deviance, however, is consistent with the observed order of acquisition in that /s/ is a late-acquiring sound and is not frequently used to substitute for the other two sounds by younger children.

Insert Table 4 about here

Acoustic Analysis

The purpose of the acoustic analysis is: 1) to provide an independent assessment of the chronology of fricative acquisition to compare with the transcription results, and 2) to investigate fricative acquisition by describing phonetic development in the two major articulatory/acoustic dimensions. To achieve both ends, adult speakers were evaluated first to demonstrate the effectiveness of selected acoustic parameters, and to serve as a baseline of comparison for children's productions. As a result, the "order of fricative acquisition" based on these analyses assumes a definition of "acquisition" to mean most similarity to the acoustic patterning in adults' productions.

Adults. A multinomial logistic regression was conducted to model fricative categorization, in which the log odds of the outcomes are modeled as a linear combination of the two acoustic predictor variables (Hosmer, & Lemeshow, 2000). In the model constructed for adult speakers, the dependent variables are the three fricative categories (/s/, /ʂ/, and /ɕ/), with /s/ being the baseline. The independent variables are normalized values of centroid frequency and onset F2 frequency. The normalization allows for direct comparison between the two variables using the coefficient to interpret their contribution to the overall model. The Wald test was used to assess statistical significance.

The results (Table 5a) illustrate that with one-unit increase in centroid frequency, the log odds for a fricative to be classified as /ʂ/ (relative to /s/) decrease by 6.889, and the log odds for a fricative to be classified as /ɕ/ (relative to /s/) decrease by 3.934. Since the baseline is set to be /s/, such decreases are in the predicted direction as /s/ has higher centroid frequency than /ʂ/ and /ɕ/. Similarly, a one-unit increment in F2 onset frequency enhances the log odds of /ʂ/ by 2.246, and increases the log odds of /ɕ/ by 4.111. Again the higher increase in the probability for /ɕ/ is expected as high F2 onset frequency is characteristic of the /ɕ/ sound. It is also interesting to note that the strength of prediction for each acoustic parameter in contributing to different contrasts. For the log odds of /ʂ/ vs. /s/, spectral mean frequency plays a more prominent role than F2 onset (an absolute value of 6.889 for centroid vs. 2.246 for F2), while the F2 onset is more important for differentiating /ɕ/ from /s/ (4.111 for F2 vs. 3.934 for centroid). All three factors included are significant. Furthermore, the model in total correctly predicts 95% of /s/, 87% of /ʂ/, and 90% of /ɕ/, demonstrating the robustness of the two acoustic parameters in describing Putonghua fricatives. Intra-rater reliability was calculated after the first author blindly transcribed 20% of the original data. This yielded a score of 98%, which demonstrated the validity of the first-pass transcription.

Insert Table 5 about here

Children. A similar multinomial analysis was conducted on child speakers. This analysis aims to quantify and predict the order of fricative acquisition by combining the two acoustic parameters. Specifically, a multinomial logistic regression model was fitted over 5-year-olds' productions first to determine the model parameters for prediction of fricative category emergence in the other age groups. In comparison with adults, the five-year-old group represents the most mature production patterns in the four age groups examined, and at the same time maintains the acoustic range more appropriate for children's speech.

The five-year-olds' model is presented in Table 5b. Similar to the adult modeling results (Table 5a), an increase in centroid frequency reduces the likelihood of /ʂ/ and /ɕ/ over /s/, whereas an increase in F2 onset enhances the likelihood of /ʂ/ and /ɕ/ relative to /s/. Furthermore, centroid frequency carries greater significance for the /ʂ/ - /s/ contrast (2.992 vs. 0.994), and F2 onset is crucial for distinguishing the /ɕ/ - /s/ contrast (3.616 vs. 1.435). Again, both parameters are statistically significant. The similar results between adults' and the 5-year-olds' model are noteworthy, as they demonstrate that five-year olds are capable of distinguishing the three fricatives in a fashion similar to adult norms, which reassures us about the suitability of classifying the rest of children's fricative productions based on the five-year-old model.

Table 6a displays the predicted production accuracy in all four age groups on the basis of the statistical model of five-year old speech. Accuracy was defined by the percentage of correct predictions in reference to the intended target. These accuracy rates agree with those from phonetic transcription in that /ɕ/ is more robustly classified based on acoustic measures than the other two sounds in all four age groups. The sound /ʂ/ ranks second, and the /s/ sound is last.

Insert Table 6 about here

Although the above-mentioned model has the advantage of providing an independent assessment to the transcription method, it assumes similar vocal tract lengths between 5-year-old children and younger-aged children or minimum impact of any differences in vocal tract length across age groups on the consequent acoustic output. Unfortunately, neither of these assumptions is valid in reality due to the fact that rapid vocal tract growth takes place during this period of life, which in turn affects the acoustic instantiation of the vocal targets (Vorperian & Kent, 2007; Vorperian et al., 2009). To avoid making these assumptions, another model was constructed on all children's data with age as a covariate to take into consideration the age-related articulatory/acoustic changes. The model was trained using native speaker's transcribed sound categories and the predictions were made according to the intended targets. Accuracies predicted by this model are presented in Table 6b. Similar to results in Table 6a, the accuracy rates agree with those from phonetic transcription in the order of /ɕ/ → /ʂ/ → /s/, particularly for the younger age groups. Therefore, although both models are limited in their own way (the five-year-old model assumes equal vocal tract length while the age-varying model relies on transcription results for training), the results converge in the same order of fricative acquisition.

It is also to be noted that, in comparing Table 2 with Table 6, the predicted accuracies based on the statistical modeling of acoustics are generally lower than the transcribed accuracies, despite the common acquisition patterns revealed by both methods. Such a discrepancy could reflect the incorporation of other acoustic cues during the process of native speaker transcription. Furthermore, native speakers are able to make auditory accommodations particularly in cases when children's speech productions were too quiet to allow for robust capturing of the two acoustic parameters examined. The fact that the first author was not blind in transcribing the data during the first pass could have also contributed to the higher accuracy of the transcription results.

The pattern of the emergence of the three fricatives in the two-dimensional acoustic space can be seen in Figure 2 (the relevant summary statistics in Table 7). At age 2, all three fricatives overlap considerably in a range of < 8000 Hz for centroid frequency and 2000~3500 Hz for F2 onset. These early forms are ambiguous between a well-formed /ʂ/ or /ɕ/ as they have the acoustic values appropriate for /ʂ/ in the centroid dimension and for /ɕ/ in the F2 dimension. By age 3, /ɕ/ starts to move to its expected acoustic region surrounding 8000 Hz in centroid frequency and around 3200 Hz for F2 onset. Overlaps remain between /ʂ/ and /s/ until age 4 when the two sounds occupy different acoustic areas, primarily differentiated from each other in the centroid dimension (<8000 Hz for /ʂ/ and > 8000 Hz for /s/). When comparing against adults, it is also evident that 5-years-olds make similar and clear distinctions between the three fricatives, but in an acoustic range higher than adults' in both dimensions. The shift in acoustic range is expected given the different vocal tract lengths between adults and children, and confirms the suitability of using five-year olds' instead of adults' speech for statistical modeling.

Insert Figure 2 about here

Insert Table 7 about here

Children's speech was further depicted for each acoustic dimension to examine the development of different aspect of articulatory motor control (Figure 3). Figure 3 regresses the acoustic values of each parameter against children's chronological age calculated in months. For each acoustic dimension, a best-fitted regression line together with a 95% confidence interval band was calculated. A lack of overlap in the confidence interval bands between the two fricatives suggests a statistically significant separation.

In the centroid dimension, the first separation starts between /ʂ/ and the other two sounds around 40 months. A second separation occurs at about 50 months between /s/ and /ɕ/. In the F2 onset dimension, a distinction between /ɕ/ and the other two sounds is already present at around 24 months, while /s/ and /ʂ/ do not further diverge from each other until after 50 months. The early distinction between /ɕ/ and /s, ʂ/ in F2 onset suggests that children are capable of raising or lowering the tongue body before 24 months. However, their control over the tongue tip to make forward or backward movement does not occur until 40 months, when /s/ separates from /ʂ/ in the centroid dimension.

Insert Figure 3 about here

Discussion

The present study combines the power of transcription with acoustic methods to determine the developmental sequence of voiceless sibilant fricatives in Putonghua-speaking children. The two measures complement each other and both agree on the order of acquisition as /ɕ/ → /ʂ/ → /s/. This order of acquisition helps to evaluate the relative contribution of the three factors proposed by previous theoretical frameworks: cross-language phoneme frequency, language-specific phoneme frequency, and children's oromotor development. Clearly, the relatively early emergence of /ɕ/ cannot be accounted for by Jakobson's phonological typology, as this sound occurs infrequently among the world's languages. By contrast, the early acquisition of /ɕ/ can be more readily attributed to the early maturation of control over the muscles that elevate the tongue body in producing this sound. However, the oromotor maturational account by itself is unable to explain why the acquisition of /s/ lags behind /ʂ/, if both involve the same articulatory muscle, the tongue tip. The sequence /ʂ/ → /s/ can be better predicted by language-specific phoneme distribution patterns, as the sound /ʂ/ occurs more often than /s/ across the Putonghua lexicon, which presumably reflects frequencies in the input to children. Thus, although neither language-specific phoneme frequency nor oral-motor development alone is able to fully account for the observed pattern, both appear to be partially correct. A theoretical model incorporating both factors but with articulatory maturation weighted more than language-specific input frequency would adequately predict the observed pattern in Putonghua fricative acquisition.

The results of the acoustic investigation are more nuanced and further illustrate in detail the developmental trajectories of children's fricative productions in each articulatory/acoustic dimension. Specifically, two-year-olds do not distinguish among the three fricatives in an adult-like manner. The sound /ɕ/ first separates from this undifferentiated acoustic space by age 3. At age 4, both /ʂ/ and /s/ occupy the acoustic range typical for /ʂ/, and the sound /s/ finally separates from /ʂ/ at age 5. It is important to note that the initial form of fricatives in 2-year-olds' speech is ambiguous between /ɕ/ and /ʂ/: it has the acoustic characteristics appropriate for /ʂ/ in the centroid dimension and appropriate for /ɕ/ in the F2 onset dimension. These acoustic characteristics correspond to a lingual gesture with a raised and retracted tongue body. This initial form with fused features of two fricatives obviously poses challenges to the auditory-based transcription analysis, and such challenges were reflected by the lower inter-rater agreement for the transcribed sounds (Table 3). Furthermore, as Li (2008) demonstrated in a perception study where 20 English-speaking adults were asked to judge the speech of children aged 2 to 5 years, the inter-listener agreement increased with increases in the child's age. All these pieces of evidence suggest the need to objectively describe children's speech using instrumental methods.

The effect of language-specific phoneme frequency is particularly evident when comparing the current study with Li (2012), which reports 100 English- and Japanese- speaking children's fricative development using similar acoustic analyses. In that study, 2-year-olds from both language backgrounds produce undifferentiated speech forms for the two fricatives, /s/ and /ʃ/, in the centroid dimension, similar to the current study. This initial merged articulation, however, was located in a frequency range at around 8000 Hz for English-speaking children, closer to a well-formed English /s/ than /ʃ/. In contrast, it was located at around 6000 Hz for Japanese-speaking children, resembling an adult-like production of Japanese /ʃ/. This language-specific difference is interpreted to reflect the distributional asymmetries of /s/ and /ʃ/ in the language-specific input: /s/ is more frequent than /ʃ/ in English, and the opposite is true for Japanese, particularly in child-directed speech (Beckman, Yoneyama, & Edwards, 2003; Chew, 1969; Tsurutani, 2004, 2007).

The robustly quantified early acquisition of /ɕ/ in Putonghua-speaking children adds to the small but diverse corpus of literature documenting the consistently early emergence of /ɕ/ in relation to other sibilant fricatives across languages (Nakanishi et al., 1972; Zharkova, 2005; Ingram, 1988b). Such cross-language consistency is unlikely to be accidental, and can be taken as an evidence for the analogous "proximal-distal principle" in tongue development. The tongue differs fundamentally from other body parts, as it is an intricately configured muscular hydrostat, containing no skeletal structures (Kier & Smith, 1985). Due to its highly complex muscular system (both intrinsic and extrinsic muscles across the sagittal, coronal, and transverse planes, Stone, 1990; 1991), the maturational mechanism of the tongue and the developmental process of children mastering motor control of different parts of the tongue are probably much more complicated than what the "proximal-distal principle" entails, and certainly warrant further investigation.

Another factor yet to be explored is the role played by children's growing perceptual capacity. In a study comparing adult perception of /s/-/ʃ/ contrast in English and Japanese, it was found that English listeners exhibit greater perceptual range for /s/, while Japanese listeners exhibit the opposite pattern (Li et al., 2011). If children display similar language-specific perceptual biases, then such biases can potentially explain the earlier acquisition of /s/ in English and /ʃ/ in Japanese. It is also to be noted that the acoustic dimension F2 onset has been shown to enjoy a perceptual advantage in English-speaking children's fricative perception development over fricative-internal cues (Nittrouer, 1992, 2002; Nittrouer & Miller, 1997). When asked to identify synthetic words beginning with /s/ and /ʃ/ where F2 onset and centroid frequency were put against each other, younger children rely more on the fricative-vowel transition cue carried by F2 onset, while older children and adults pay greater attention to the fricative-internal cue (i.e., centroid frequency). If such a perceptual bias towards F2 onset is universally valid, it provides an alternative account to the oromotor maturation hypothesis for the early acquisition of /ɕ/ in Mandarin-speaking children. Studies are needed to determine whether such a perceptual bias can also be found in Mandarin-speaking children.

One last caveat is that neither of two cited corpora contains Putonghua frequency calculated based on child-directed speech. The Lancaster Corpus of Mandarin Chinese is compiled on written texts. Although the CALLHOME Mandarin lexicon reports phoneme frequency of spoken language, the data collected were not addressed primarily to children.

Future research will need to confirm that these frequency patterns hold in the input that children directly receive.

Finally, although the current study is primarily designed to address theoretical questions, it has implications for clinicians dealing with Putonghua-speaking individuals with communication disorders. Speech-language pathology as a profession is relatively new in China. No diagnosis of speech sound disorder can be made from observations of a small subset of sounds in a language. However, based on our study, we would predict that errors in /ɕ/ would be less among Putonghua children than would errors /s/ and /ʂ/, as /ɕ/ is the earliest acquired sound and thus is predicted to be most accurate. We would also predict errors of /s/ to be the most common of all three of these fricatives. Finally, if our interpretation of the acoustic analysis is correct, then we predict that the accurate production of /ɕ/ would be related to the ability to maneuver the tongue dorsum in the coronal plane of the oral cavity. Although such a prediction awaits to be verified from future direct imaging studies, results of our study suggest the potential use of /ɕ/ as a diagnostic marker for speech pathology and points to the possible physiological basis of any potential delay associated with /ɕ/ production.

Acknowledgment

Data collection and analysis were supported by the Ohio State University Target Investment Fellowship to Fangfang Li and Eunjong Kong, NIDCD grant 02932 to Jan Edwards, and University of Lethbridge Start-up Fund to Fangfang Li. We thank the staff in Songyuan No. 2 Daycare Center in facilitating participant recruitment and testing. We also thank those children and adults who participated in the study. Further thanks go to Mary E. Beckman and Jan Edwards for their advice in designing the experiment and transcription protocol, and for extremely useful input in the development of the acoustic analysis employed in this study, and to Jennifer Mather for help with proofreading this article.

References

Baum, S. R., & McNutt, J. C. (1990). An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics, 18*, 51-63.

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

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

Boruta, L., & Jastrzebska, J. (2012). *A phonemic corpus of Polish child-directed speech.* Paper presented at the Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey.

Bosma, J. F. (1975). Anatomic and physiologic development of the speech apparatus. In D. B. Towers (Ed.), *Human communication and its disorders* (pp. 469-481). New York: Raven.

Butterworth, G., Verweij, E., & Hopkins, B. (1997). The development of prehension in infants: Halverson revisited. *British Journal of Developmental Psychology, 15*, 223–236.

Chevrie-Muller, C., & Lebreton, M. T. (1973). [Study of the pronunciation of consonants during a word repetition test in groups of children aged 3 and 5 and one-half years]. *Rev Laryngol Otol Rhinol (Bord), 94*(3), 109-152.

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

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

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of

British English-speaking children. *Clinical Linguistics & Phonetics, 17*(8), 617-643.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying phonological acquisition. *Clinical Linguistics and Phonetics, 22*(12), 939-958.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51*(2), 419-439.

Fletcher, S. G. (1989). Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech and Hearing Research, 32*, 736-748.

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.

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 postalveloar fricatives of Polish. In K. Shigeru, H. Hirose & H. Fujisaki (Eds.), *Speech Production and Language: In Honor of Osamu Fujimura* (Vol. 13, pp. 176-191). New York: Mouton de Gruyter.

Hosmer, D., & Lemeshow, S. (2000). Applied Logistic Regression (pp. 31-46). 2nd edition. New York (NY): John Wiley & Sons, Inc

Hu, F. (2008). The three sibilants in Standard Chinese. In S. F. R.Sock, & Y. Laprie (Ed.), *Proceedings of the 8th International Seminar on Speech Production (ISSP 08)* (pp. 105-108). Strasbourg, France: INRIA.

Huang, S., Bian, X., Wu, G., & McLemore, C. (1997). LDC Mandarin Lexicon. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Ingram, D. (1988a). The acquisition of word-initial [v]. *Language and Speech, 31*, 77-85.

Ingram, D. (1988b). Jakobson revisited: Some evidence from the acquisition of Polish. *Lingua, 75*, 55-82.

Ingram, D., Christensen, L., Veach, S., & Webster, B. (1980). The acquisition of word-initial fricatives and affricates in English by children between 2 and 6 years. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), *Child Phonology* (Vol. 1. Production, pp. 169-192). New York: Academic Press.

Irwin, O. C. P. (1933). Proximodistal differentiation of limbs in young organisms. *Psychological Review, 40*, 467-477.

Jakobson, R. (1941/1968). *Child Language, aphasia, and phonological universal.* Mouton: The Hague.

Jeng, H. (1979). The acquisition of Chinese phonology in relation to Jakobson's laws of irreversible solidarity *Proceedings of the 9th International Congress of Phonetic Sciences* (Vol. 2, pp. 155-161): University of Copenhagen.

Kent, R. D. (1983). The segmental organization of speech. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 57-89). New York: Springer-Verlag.

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*(3), 7-23.

Kier, W., & Smith, K. (1985). Tongue, tenacles, and trunks: The biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnaean Society, 83*, 307-324.

Hu, F. (2008). The three sibilants in Standard Chinese *8th International Seminar on Speech Production* (pp. 105-108). Strasbourg, France.

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

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108*(3), 1252-1263.

Ladd, D.R. (2011). 2011. Phonetics in phonology. In J. Goldsmith, J. Riggle, and A. Yu (Eds.), *Handbook of Phonological Theory, 2nd Edition* (p. 348-373). New York: Blackwell.

Ladefoged, P., & Wu, Z. (1984). Places of articulation: An investigation of Pekingese fricatives and affricatives. *Journal of Phonetics, 12*(3), 267-278.

Lee, W. (1999). An articulatory and acoustic analysis of the syllable-initial sibilants and approximants in Beijing Mandarin *Proceedings of the 14th International Congress on Phonetic Sciences (ICPhS 99)* (pp. 413-416). San Francisco, USA.

Li, C. N., & Thompson, S. A., 185±99. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 185-199.

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese, and Mandarin Chinese.* Ph. D. Thesis, Ohio State University, Columbus, Ohio.

Li, F. (2012). Language-specific developmental differences in speech production: A cross-language acoustic study. *Child Development, 83*(4), 1303-1315.

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, W., Zhu, H., & Dodd, B. (2002). Phonological saliency and phonological acquisition by Putonghua speaking children: A cross-populational study. In F. Windsor, M. L. Kelly & N. Hewlett (Eds.), *Investigations in Clinical Phonetics and Linguistics* (pp. 169-184). Mahwah, NJ: Erlbaum.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61*, 379–388.

Lin, T., & Wang, L. J. (1992). 语音学教程 *(A Course in Mandarin Phonetics)*. Beijing: Peking

University Press.

Lin, Y. S., & Peng, S. C. (2003). Acquisition profiles of syllable-initial consonants in Mandarin-speaking children with cochlear implants. *Acta Otolaryngol, 123*(9), 1046-1053.

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.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

McBryde, C., & Ziviani, J. (1990). Proximal and distal upper limb motor development in 24 week old infants. *Canadian Journal of Occupational Therapy, 57*(3), 147-154.

McEnery, A. M., & Xiao, Z. (2004). The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC) 2004, 17*, 1175-1178.

McGowan, R. S., & Nittrouer, S. (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/. *Journal of the Acoustical Society of America, 83*(1), 229-236.

Munson, B., Johnson, J., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison of speech-language pathologists with clinically untrained listeners. *American Journal of Speech-Language Pathology, 24*, 124-139.

Munson, B., Edwards, J., Schellinger, S.K., Beckman, M.E., & Meyer, M.K. (2010). Deconstructing phonetic transcription: covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. *Clinical Linguistics and Phonetics, 24*, 245-260.

Nakanishi, Y., Owada, K., & N., F. (1972). Kōon kensa to sono kekka ni kansuru kōsatu. . *Tokushū kyōiku kenkyū shisetu hōkoku (Bulletin of the Tokyo Gakugei University Special Education*

Research Group) 1, 1-41.

Nguyen, N., Marchal, A., & Content, A. (1996). Modeling tongue-palate contact patterns in the production of speech. *Journal of Phonetics, 24*, 77-97.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics, 20*(3), 351-382.

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*(1), 520-530.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America, 112*(2), 711-719.

Nittrouer, S, & Miller, M. E. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America, 102*(1), 572-580.

Noiray, A., Menard, L., & Iskarous, K. (2013). The development of motor synergies in children: ultrasound and acoustic measurements. *Journal of the Acoustical Society of America, 133*(1), 444-452.

Poole, I. (1934). Genetic development of articulation of consonants sounds in speech. [Large normal study of English children's articulation]. *Elementary English Review, 11*, 159-161.

Prather, E., Hedrick, D., & Kern, D. (1975). Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders, 53*, 12.

Pye, C., Ingram, D., & List, H. (1987). A comparison of initial consonant acquisition in English and Quiché. In K. E. Nelson & A. V. Kleeck (Eds.), *Children's Language* (Vol. 6, pp. 175-190). Hillsdale, NJ: Lawrence Erlbaum.

R Development Core Team (2011). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rahim, K. (2010). Multitaper: Mutitaper Spectral Analaysis. (Version R package version 0.1-2). Retrieved from http://CRAN.R-project.org/package=multitaper

Recasens, D. (2013). On the articulatory classification of (alveolo)palatal consonants. *Journal of the International Phonetic Association, 43*(1), 1-22.

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37*, 55-63.

Scobbie, J., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: language Acquisition and the Lexicon* (pp. 194-203). Cambridge: Cambridge University Press.

Shih, Y., & Kong, E. J. (Eds.). (2011). *Perception of Mandarin fricatives by native speakers of Taiwan Mandarin and Taiwanese* (Vol. 1). University of Oregon, Eugene.

Si, Y. (2006). 普通话儿童语音习得的个案研究 (A case study of Putonghua-speaking children's phonetic development). 当代语言学 *(Modern Linguistics), 1*, 1-16.

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.

Stevens, K. N., Li, Z., Lee, C., & Keyser, S. J. (2004). A note on Mandarin fricatives and enhancement. In G. Fant, H. Fujisaki, J. Cao & Y. Xu (Eds.), *From traditional phonology to modern speech processing* (pp. 393-403). Beijing: Foreign Language Teaching and Research Press

Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and X-ray microbeam data. *Journal of the Acoustical Society of America, 87*(5), 2207-2217.

Stone, M. (1991). Toward a three-dimensional model of tongue movement. *Journal of Phonetics, 19*, 309-320.

Studebaker, G. A. (1985). A "Rationalized" Arcsine Transform. *Journal of Speech, Language, and Hearing Research, 28*, 455-462.

Svantesson, J. (1986). Acoustic analysis of Chinese fricatives and affricatives. *Journal of Chinese Linguistics, 14*(1), 53-70.

Templin, M. (1957). *Certain Language Skills in Children.* (Vol. 26). Minneapolis: Univ. Minnesota.

Toda, M., & Honda, K. (2003). *An MRI-based cross-linguistic study of sibilant fricatives.* Paper presented at the Proceedings of the 6th International Seminar on Speech Production, Sydney, Australia.

Tsurutani, C. (2004). Acquisition of Yo-on (Japanese contracted sounds) in L1 and L2 phonology in Japanese second language acquisition. *Journal of Second Language, 3*, 21.

Tsurutani, C. (2007). Early acquisition of palato-alveolar consonants in Japanese: Phoneme frequencies in child-directed speech. *Journal of the Phonetic Society of Japan, 11*(1), 102-110.

Umebayashi, A., & Takagi, S. (1965). Gakureimae no kodomo no koonnoryoku ni kansuru ichikenkyu (Articulatory competence of children under school age). *Onseigengo Igaku, 6*, 17-18.

Vorperian, Houri K., & Kent, Ray D. (2007). Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. *Journal of Speech, Language and Hearing Research, 50*, 1510-1545.

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of the Acoustical Society of America, 125*(3), 1666-1678.

Wallace, P. S., & Whishaw, I. Q. (2003) Independent digit movements and precision grip patterns in 1-5-month-old human infants: Hand-babbling, including vacuous then self-directed hand and digit movements, precedes targeted reaching. *Neuropsychologia, 41*(14), 1912-1918.

Wellman, B., Case, I., Mengert, I., & Bradbury, D. (1931). Speech sounds of young children. *University of Iowa Studies in Child Welfare, 5*, 1–82.

Xu, L., Yang, W., & Qi, G. (2010). 汉语学龄前儿童普通话辅音音位习得的自然音系学分析 - 名两周岁十一个月儿童的个案研究 (A natural phonology analysis of Putonghua consonant acquisition in Chinese Preschool children: A case study of a two year old eleven month child). 宁波大学学报*(Journal of Ningbo University), 23*(2), 64-68.

Zharkova, N. (2005). Strategies in the acquisition of segments and syllables in Russian-speaking children. In M. Tzakosta, C. Levelt & J. v. d. Weijer (Eds.), *Developmental Paths in Phonological Acquisition. Special issue of Leiden Papers in Linguistics, 2.1* (pp. 189-213).

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language, 27*, 3-42.

Appendix A

The word list

target vowel	*/s/*	*Chinese orthography*	*/ɕ/*	*Chinese orthography*	*/ʂ/*	*Chinese orthography*
/a/	/sa.niao/	撒尿	/ɕa.jy/	下雨	/ʂa.fa/	沙发
	/san/	伞	/ɕaŋ.tsi/	箱子	/ʂan/	山
	/san.tɕiao/	三角	/ɕaŋ.tɕao/	香蕉	/ʂan.yaŋ/	山羊
	/sa.lə/	撒了	/ɕa/	虾	/ʂan.tsi/	扇子
/i/	/si.tɕi/	司机	/ɕi. ʂo/	洗手	/ʂi.tsi/	石子
	/si.kə/	四个	/ɕi/	细	/ʂi/	十
	/si/	丝	/ɕi.kua/	西瓜	/ʂi.tsi/	狮子
	/si.miao/	寺庙	/ɕiŋ. ɕiŋ/	星星	/ʂi.ts/	柿子
/o/	/soŋ.ʂu/	松鼠	/ɕo. ɕi/	休息	/ʂo.tɕyan/	手绢
	/soŋ/	松	/ɕo/	袖	/ʂo.tʰao/	手套
	/soŋ.tɕin.tai/	松紧带	/ɕoŋ.mao/	熊猫	/ʂo.tsi/	瘦子
	/soŋ.lə/	松了	/ɕoŋ/	熊	/ʂo/	手
/u/	/sun/	笋			/ʂu.pao/	书包
	/su.liao/	塑料			/ʂu. ʂu/	叔叔
	/sun.tsi/	孙子			/ʂu/	树
	/sun.wu.koŋ/	孙悟空			/ʂu.tsʰai/	蔬菜
/ɛ/			/ɕɛn.hua/	鲜花		

	/ɕɛn/	线
	/ɕɛ.tsi/	写字
	/ɕɛ.tsi/	鞋子

Table 1. Age and gender breakdown of participants with the mean and standard deviation of the age (in months) specified for each group.

	2-year-olds (M=30 months SD= 3 months)	*3-year-olds (M=42 months SD= 3 months)*	*4 –year-olds (M=54 months SD= 4 months)*	*5-year-olds (M=66 months SD= 3 months)*	*Adults (18-30) (M=270 months SD= 31 months)*
Females	12	13	14	12	10
Males	12	12	9	13	10
Total	24	25	23	25	20

Table 2. Number of tokens (proportions) of /s/, /ʂ/, /ɕ/ that were judged to be correct, segregated by vowel context and children's age group.

Age	*Vowel*	*/s/*		*/ʂ/*		*/ɕ/*	
		Total Number	*Correct number (proportion)*	*Total number*	*Correct number (proportion)*	*Total number*	*Correct number (proportion)*
2 (n=1139)	/a/	96	21 (.22)	94	43 (.46)	95	70 (.74)
	/ɛ/	0	*NA*	0	*NA*	97	60 (.62)
	/i/	94	14 (.15)	92	43 (.47)	94	63 (.67)
	/o/	96	20 (.21)	93	44 (.47)	96	53 (.55)
	/u/	94	16 (.17)	98	42 (.43)	0	*NA*
	Overall	380	71 (.19)	377	173 (.46)	382	246 (.64)
3 (n=1199)	/a/	100	50 (.50)	100	73 (.73)	100	91 (.91)
	/ɛ/	0	*NA*	0	*NA*	99	92 (.93)
	/i/	100	53 (.53)	100	76 (.76)	100	95 (.95)
	/o/	100	49 (.49)	100	72 (.72)	100	89 (.89)
	/u/	100	48 (.48)	100	75 (.75)	0	*NA*
	Overall	400	200 (.50)	400	296 (.74)	399	367 (.92)
4 (n=1090)	/a/	93	81 (.87)	92	79 (.86)	91	88 (.97)
	/ɛ/	0	*NA*	0	*NA*	90	89 (.99)
	/i/	89	79 (.89)	91	81 (.89)	92	92 (1.00)
	/o/	89	67 (.75)	91	82 (.90)	90	87 (.97)
	/u/	91	70 (.77)	91	84 (.92)	0	*NA*
	Overall	362	297 (.82)	365	326 (.89)	363	356 (.98)
5 (n=1197)	/a/	100	90 (.90)	100	93 (.93)	99	98 (.99)
	/ɛ/	0	*NA*	0	*NA*	100	99 (.99)
	/i/	100	93 (.93)	100	95 (.95)	100	99 (.99)
	/o/	99	86 (.87)	100	95 (.95)	100	97 (.97)
	/u/	99	89 (.90)	100	93 (.93)	0	*NA*
	Overall	398	358 (.90)	400	376 (.94)	399	393 (.98)
Total		1540	926 (.60)	1542	1171 (.76)	1543	1362 (.88)

Table 3. Number (percentage) of the transcribed sounds for the 614 tokens of /s/, 371 tokens of /ʂ/, and 181 tokens of /ɕ/ that were judged to be incorrect. Note that only error patterns that represent more than 1% of the total errors for each target are reported. In addition, percentage of agreement between the two transcribers was noted inside the brackets for each transcribed error pattern whenever applicable.

Target	*Fricative place*		*Stopping*		*Affrication*		*Deletion*
Fricative	*Error*	*Number (percentage)*	*Error*	*Number (percentage)*	*Error*	*Number (percentage)*	*Number (percentage)*
/s/	[ʂ]	331(54%)[81%]	[t^{h}]	9 (1%)	[tʃ]	17 (3%)	17 (3%)
	[ɕ]	86 (14%)[84%]	[p^{h}]	7 (1%)	[tɕ^{h}]	14 (2%)	
	[θ]	55 (9%)	[k]	6 (1%)	[ts^{h}]	9 (1%)	
	[f]	6 (1%)	[t]	5 (1%)	[ts]	7 (1%)	
/ʂ/	[s]	104 (28%)[84%]	[t^{h}]	9 (2%)	[tʃ]	45 (12%)	22 (6%)
	[ɕ]	74 (20%)[70%]	[t]	2 (1%)	[tʃ^{h}]	17 (5%)	
	[h]	43 (11%)[67%]			[tɕ^{h}]	4 (1%)	
	[f]	14 (4%)			[ts]	3 (1%)	
	[θ]	7 (2%)					
/ɕ/	[ʂ]	72 (40%)[62%]	[t]	5 (3%)	[tɕ^{h}]	29 (16%)[57%]	9 (5%)
	[s]	14 (8%)	[t^{h}]	4 (2%)	[tɕ]	11 (6%)	
	[h]	11 (6%)	[k^{h}]	2 (1%)	[tʃ]	7 (4%)	
	[θ]	4 (2%)					

Table 4. The age breakdown of the distribution of the top three error types for each target fricative, as assessed by native-speaker transcription over all children.

Target fricative	*Error types*	*Total number of errors*	*Age group*			
			2	3	4	5
/s/	[ʂ]	331	147	141	37	6
	[ɕ]	86	68	18	0	0
	[θ]	55	37	18	0	0
/ʂ/	[s]	104	19	40	26	19
	[ɕ]	74	59	11	0	4
	[tʃ]	45	30	15	0	0
/ɕ/	[ʂ]	72	60	7	0	5
	[tɕʰ]	29	20	9	0	0
	[s]	14	5	5	4	0

Table 5. Coefficients of the multinomial logistic regression model conducted for adults (a) and children (b). In each model, the log odds of the outcomes (three levels, /s/ vs. /ʂ/ vs. /ɕ/) are modeled as a linear combination of the predictor variables. The dependent variables were the three target fricatives (/s/ as the baseline) and the independent variables were standardized values of centroid frequency and F2 onset.

(a) Adults:

		Coefficient	*Std. error*	*z value*	*p value*
/ʂ/ vs. /s/ (baseline)	Intercept	0.496	0.338	1.471	< 0.001
	Centroid frequency	-6.889	0.504	-13.661	< 0.001
	Onset F2 frequency	2.246	0.363	6.178	< 0.001
/ɕ/ vs. /s/ (baseline)	Intercept	1.740	0.296	5.873	< 0.001
	Centroid frequency	-3.934	0.432	-9.091	< 0.001
	Onset F2 frequency	4.1097	0.357	11.491	< 0.001
Residual Deviance 626.696; AIC: 638.696					

(b) Children:

		Coefficient	*Std. error*	*z value*	*p value*
/ʂ/ vs. /s/ (baseline)	Intercept	0.152	0.148	1.028	< 0.001
	Centroid frequency	-2.992	0.193	-15.540	< 0.001
	Onset F2 frequency	0.994	0.161	6.179	< 0.001
/ɕ/ vs. /s/ (baseline)	Intercept	-0.051	0.158	-0.318	< 0.001
	Centroid frequency	-1.435	0.184	-7.799	< 0.001
	Onset F2 frequency	3.616	0.219	16.441	< 0.001
Residual Deviance: 1045.497; AIC: 1057.497					

Table 6: Percentage of predicted accuracy for each fricative produced by children aged 2 to 5 years, on the basis of two multinomial models. In (a), predictions were made by fitting all children's data over the model constructed over 5-year-old's productions. The model has the dependent variables being the intended/canonical fricative categories, and the independent variables being centroid frequency and onset F2 frequency values of the five-year old productions. In (b), predictions were made based on the model constructed on the data from children in all four age groups. The dependent variables were the transcribed fricative categories, and the independent variables are the centroid frequency and onset F2 frequency values of all children's productions. Children's age was coded as a covariate in (b).

(a)

	/s/	/ʂ/	/ɕ/
2-year-olds	12%	40%	80%
3-year-olds	34%	67%	94%
4-year-olds	73%	70%	95%
5-year-olds	83%	82%	88%

(b)

	/s/	/ʂ/	/ɕ/
2-year-olds	11%	53%	70%
3 year olds	42%	73%	85%
4-year-olds	73%	78%	89%
5-year-olds	84%	89%	88%

Table 7: Summary statistics (mean and standard deviation) of the two acoustic parameters for children of the four age groups as well as for adults.

	Centroid frequency (Hz)			F2 onset frequency (Hz)		
	/s/	/ʂ/	/ɕ/	/s/	/ʂ/	/ɕ/
2-year-olds	6520.91	6340.41	6504.46	2566.35	2543.41	2992.71
	(1837.66)	(2055.84)	(1597.65)	(555.00)	(571.78)	(424.31)
3-year-olds	7299.66	5877.58	7333.70	2298.68	2359.00	3084.60
	(2238.00)	(1474.07)	(1416.90)	(458.38)	(449.74)	(317.92)
4-year-olds	9216.41	5902.89	8175.19	2150.47	2268.74	3014.85
	(2279.96)	(1512.62)	(1632.53)	(398.66)	(446.14)	(271.53)
5-year-olds	9260.14	5565.52	7651.34	2002.25	2167.11	2938.35
	(1909.11)	(1330.85)	(1453.62)	(347.06)	(419.41)	(327.18)
Adults	8342.36	4758.99	6495.64	1470.32	1682.09	2174.75
	(1216.25)	(880.03)	(1147.89)	(241.54)	(327.41)	(291.36)

Figure captions:

Figure 1: Transcribed accuracy of the three fricatives produced by children as a function of their age. /s/ is represented by red square, /ʂ/ is represented by blue dot, and /ɕ/ is represented by green triangle.

Figure 2: Scatterplot of fricative productions by adults and children in each age group, plotted according to centroid and F2 onset frequency. /s/, /ʂ/, and /ɕ/ are represented by red square, blue circle, and green triangle, respectively.

Figure 3: Scatterplot of averaged centroid and F2 onset values for each fricative token produced by children as a function of their age (in months). Each child was represented by one symbol. Straight lines are the best-fitted lines for each fricative, with the dependent variable being the acoustic value and the independent variable being children's age. The curved lines represent the 95% confidence interval of the best-fitted line. /s/, /ʂ/, and /ɕ/ are represented by red circle, green plus, and blue triangle, respectively.

Figure 1.

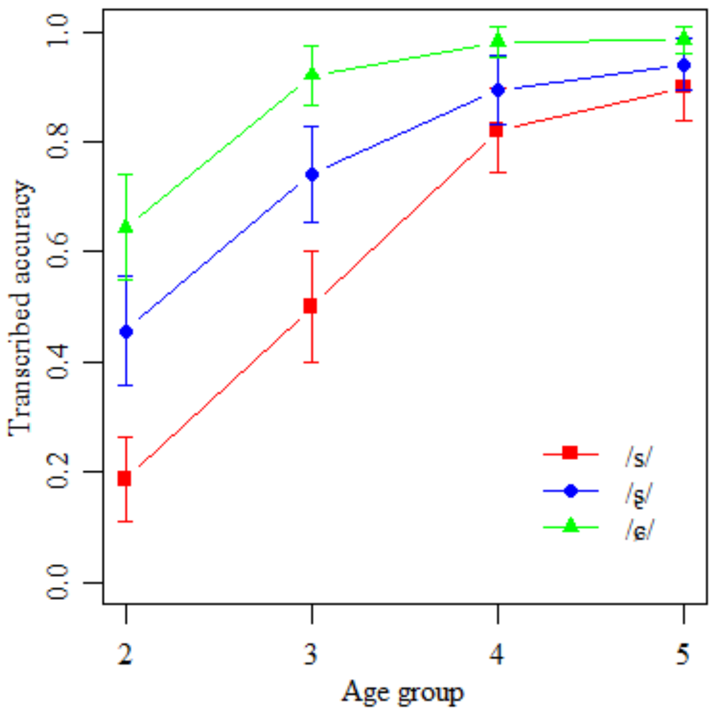


Figure 2.

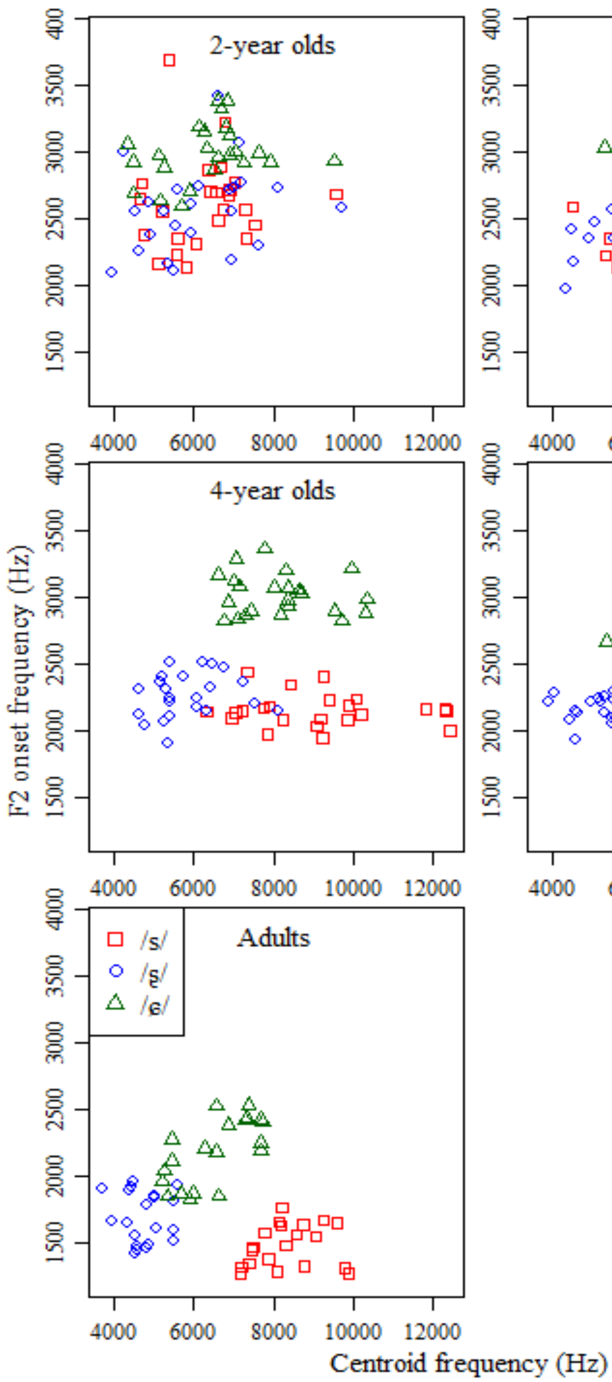
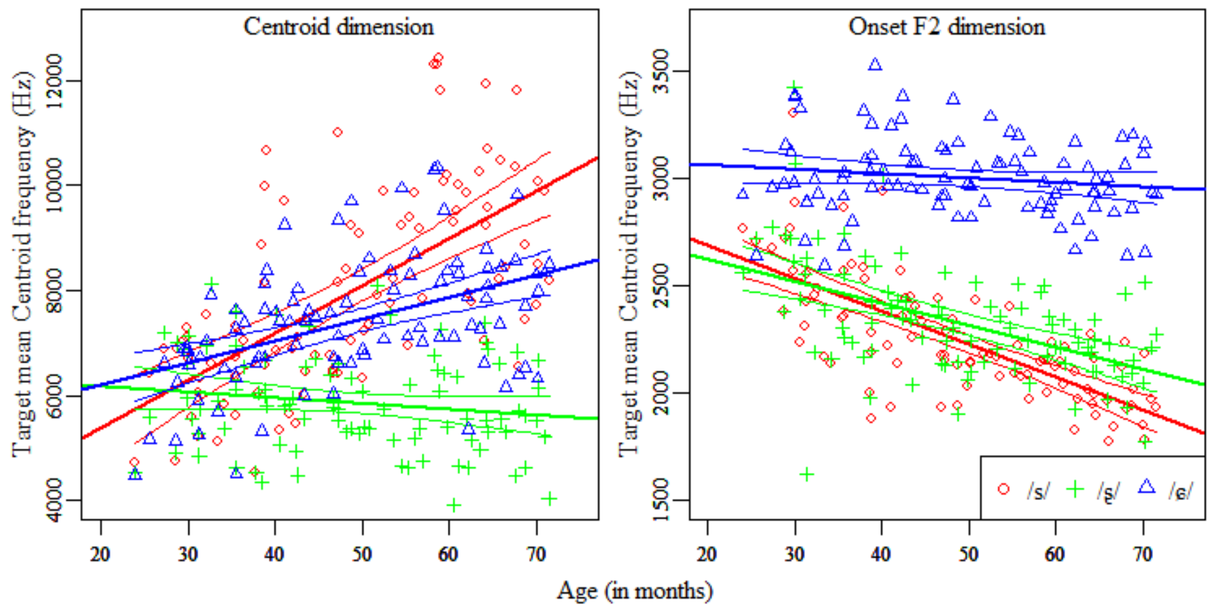

2-year olds
3-year olds
4-year olds
5-year olds
Adults
/s/
/ʂ/
/ɕ/
F2 onset frequency (Hz)
Centroid frequency (Hz)
1500
2000
2500
3000
3500
4000
4000
6000
8000
10000
12000

Figure 3.


Centroid dimension
Onset F2 dimension
Target mean Centroid frequency (Hz)
Target mean Centroid frequency (Hz)
4000
6000
8000
10000
12000
1500
2000
2500
3000
3500
20
30
40
50
60
70
Age (in months)
/s/
/ʂ/
/ɕ/