Research Article

The Development of Voiceless Sibilant Fricatives in Putonghua-Speaking Children

Fangfang Li^a and Benjamin Munson^b

Purpose: The aims of the present study are (a) to quantify the developmental sequence of fricative mastery in Putonghua-speaking children and discuss the observed pattern in relation to existing theoretical positions, and (b) to describe the acquisition of the fine-articulatory/acoustic details of fricatives in the multidimensional acoustic space.

Method: Twenty adults and 97 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 used to index tongue-body and tongue-tip development, respectively.

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

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; Ferguson & Farwell, 1975; Ingram, 1988a; Jakobson, 1941/1968; Kent, 1992; F. Li, 2012; Locke, 1983; Pye, Ingram, & List, 1987). 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 that is 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 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 with 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 cross-language studies of phonological acquisition, suggesting the importance of phonemes’ language-specific frequency of occurrence (Ingram, 1988a; Pye et al., 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 et al.,

^aUniversity of Lethbridge, Alberta, Canada

^bUniversity of Minnesota, Minneapolis

Correspondence to Fangfang Li: fangfang.li@uleth.ca

Editor: Jody Kreiman

Associate Editor: Ben A. M. Maassen

Received May 27, 2014

Revision received January 22, 2015

Accepted January 2, 2016

DOI: 10.1044/2016_JSLHR-S-14-0142

Disclosure: The authors have declared that no competing interests existed at the time of publication.

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 (Ladefoged & Maddieson, 1996; Maddieson, 1984). Pye et al. 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 the UCLA Phonological Segment Inventory Database for references on phoneme frequencies is that it is 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 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 (Gibbon, 1999; F. Li, 2008). Such a hypothesis could readily explain a peculiar acquisition phenomenon concerning the alveolopalatal fricative /ç/. This sound is often reported to emerge earlier than other voiceless sibilant fricatives in children's speech in many languages (Nakanishi, Owada, & Fujita, 1972, for Japanese; Zharkova, 2005, for Russian; Ingram, 1988b, for Polish), despite its rarity both across languages and within some languages. The main difference between /ç/ and other sibilant fricatives lies in its utilization of the large, more central muscles controlling the dorsum as the major articulator as compared with the use of the smaller, more peripheral muscles that control the tongue tip in /s/ or /ʃ/.

Methodological Challenges

All these theoretical controversies regarding 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, Edwards, Schellinger, Beckman, &

Meyer, 2010; Munson, Johnson, & Edwards, 2012; 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; F. Li, Edwards, & Beckman, 2009; Macken & Barton, 1980; Scobbie et al., 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." This assumption has not been upheld by instrumental studies. The developmental process of speech sounds revealed by instrumental analysis is more gradual than transcription studies have suggested (F. Li, 2012; Macken & Barton, 1980; Nittrouer, 1995; Scobbie et al., 2000).

On the other hand, although instrumental analysis offers rich information on children's speech profiles, 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 purposes of the present study are twofold. The first is to solidify the chronology of the 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 use 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. Second, 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 because it contains a rich inventory of three voiceless sibilant fricatives: alveolar/dental /s/, alveolopalatal /ç/, and retroflex /ʂ/ (Ladefoged & Wu, 1984; T. 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 is termed a *flat retroflex* (Ladefoged

& Maddieson, 1996; Ladefoged & Wu, 1984). The following sections review the literature on Putonghua fricatives and make predictions about the Putonghua-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 UCLA Phonological Segment Inventory Database reveals the prevalence of /s/, occurring in 197 languages, in contrast to the rarity of /ç/ (21 languages) and /ʃ/ (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. This yielded the highest frequency for /ʃ/ (32,357 times), followed by /ç/ (23,199 times) and lastly /s/ (6,641 times). The Lancaster Corpus is a Chinese corpus sampling 15 written text categories such as news, literary texts, academic prose, and official documents, published in mainland China around 1991, with a total of approximately one million words (McEnery & Xiao, 2004).

Similar findings were obtained by F. Li (2008), who reported the log frequency of voiceless sibilant fricatives in Putonghua using the CALLHOME Mandarin Chinese lexicon retrieved from the online Linguistic Data Consortium catalog (Huang, Bian, Wu, & McLemore, 1997). The corpus consists of 44,405 words of conversational speech. F. Li 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; reported log frequency is highest for /ʃ/ (−2.7), followed by /ç/ (−2.8) and /s/ (−3.9). On the basis of these two studies, if language-specific phoneme frequency plays a decisive role in determining the acquisition order of fricatives, /ʃ/ and then /ç/ will be mastered earlier than /s/.

Articulation Characteristics of Putonghua Fricatives

Both /s/ and /ʃ/ are articulated with the tongue apex (Hu, 2008; Ladefoged & Wu, 1984; Lee, 1999). The distinction between /s/ and /ʃ/ 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, whereas that for /ʃ/ is posterior to the alveolar ridge (Hu, 2008; Ladefoged & Wu, 1984; Lee, 1999). In contrast, it is the tongue dorsum instead of the apex that is used for the articulation of /ç/; specifically, the tongue predorsum is elevated toward the hard palate. This creates a long palatal passage to channel forced air (Ladefoged & Wu, 1984; T. 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 /ç/ should be acquired prior to control over those that are used to produce /s/ and /ʃ/. If tongue maturation is the primary or only factor in determining Putonghua fricative acquisition order, then /ç/ would be expected to precede the sounds /s/ and /ʃ/.

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; C. Li & Thompson, 1977; Y. 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 /ʃ/, as Taiwanese does not have the /ʃ/ sound (Shih & Kong, 2011).

The handful of studies examining the development of Putonghua in children have yielded mixed results regarding the developmental sequence of fricatives (i.e., W. Li, Zhu, & Dodd, 2002; Si, 2006; Xu, Yang, & Qi, 2010). In a longitudinal study investigating the speech development of four Putonghua-speaking children, W. Li et al. (2002) found that two children produced /ç/ first, whereas the other two children produced /s/ first. Zhu and Dodd (2000) conducted a large-scale cross-sectional study of 129 children aged 1 to 4 years. The study revealed an earlier mastery of /ç/ (2;7 [years;months]–3;0), followed by /s/ (4;1–4;6), with /ʃ/ (>4;6) acquired last. However, because that study targeted the entire Putonghua consonant inventory, the sampling of the three voiceless sibilant fricatives was both sparse and unbalanced: /ç/ and /ʃ/ were elicited eight and four times, respectively, in word-initial position, with /s/ 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 that have proven effective in previous research on Mandarin Chinese were applied in the acoustic analysis of the present study: (a) centroid frequency, calculated from the middle portion of fricative noise; and (b) 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 toward a more anterior position in the oral cavity (Jongman, Wayland, & Wong, 2000; McGowan & Nittrouer, 1988; Shadle & Mair, 1996). As a consequence, the centroid frequency of Mandarin fricatives is expected to vary in the order /s/ > /ç/ > /ʃ/ (F. Li, 2008; Svantesson, 1986). The centroid measure can index the location of the lingual constriction forward or backward in the midsagittal plane of the oral cavity.

The second measure, F2 onset frequency, is inversely correlated with the length of the back cavity, and /ç/ is expected to exhibit the highest value due to its distinct tongue posture (Halle & Stevens, 1997; F. Li, 2008; Stevens, Li, Lee, & Keyser, 2004). Unlike /s/ or /ʃ/, 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 toward the hard palate in producing /ç/. As a result of this posture, the back cavity of /ç/ is much reduced in comparison to that of /s/ or /ʃ/. Between /s/ and /ʃ/, the length of the back cavity is slightly longer for /ʃ/, owing to its retracted tongue-tip position, but the length of the back cavity is longer for both sounds than for /ç/. 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.

Method

Participants

Ninety-seven children aged 2;0–5;0 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 typical hearing and passed a hearing screening using otoacoustic emissions at 2000, 3000, 4000, and 5000 Hz. No child tested had any reported speech, language, or hearing problems, according to parent or teacher reports. 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.

Task and Materials

Children were tested individually in a quiet room of a day-care 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 first listening to the word the computer says and then repeating 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 lasted approximately 5–10 min. 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 with children, except that they were informed that the experiment was 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 for the entire word list). Due to phonotactic constraints, no words were included for the sequences */se/, */se/, and */cu/. Thus, a total of 4,656 tokens (16 × 3 targets × 97 children) were elicited from children. However, due to circumstances such as productions with unintended or unrecognizable words, 31 tokens were removed from the transcription analysis, yielding a total of 4,625 tokens (1,540 for /s/, 1,542 for /ʃ/, and 1,543 for /ç/). 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 4,240 remaining tokens for children. For adults, 960 tokens (16 × 3 targets × 20 adults) were elicited and included in the acoustic analysis.

Procedure

Participants' 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, a total of 4,625 tokens were transcribed and included in the statistical analysis. The first author, a native speaker of Putonghua

Table 1. Age and gender breakdown of participants.

Gender	2 years ^a	3 years ^b	4 years ^c	5 years ^d	Adults (18–30 years) ^e
Female	12	13	14	12	10
Male	12	12	9	13	10
Total	24	25	23	25	20

^aM = 30 months, SD = 3 months. ^bM = 42 months, SD = 3 months.

^cM = 54 months, SD = 4 months. ^dM = 66 months, SD = 3 months.

^eM = 270 months, SD = 31 months.

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 (four 2-year-olds, five 3-year-olds, five 4-year-olds, and five 5-year-olds), and the phoneme-by-phoneme interrater reliability was 85% for /ç/, 89% for /ʃ/, 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 the first zero crossing in the upswing voicing cycle of the following vowel. Centroid frequency was calculated from a Multitaper spectrum on the basis of the middle 40-ms slice of each fricative segment using the Multitaper package (see Appendix A in Rahim, Burr, & Thomson, 2014) in R software (R Core Team, 2011). Each Multitaper spectrum was high-pass filtered (above 1000 Hz) to eliminate potential low-frequency noise such as blowing wind or an opening or closing door.

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

/ç/ was slightly more accurate than /ʃ/. Both /ç/ and /ʃ/ show much higher accuracy rates than /s/. Around 4 years old, the gap between /s/ and the other sound starts to close, and by 5 years old, all three sounds have rates higher than .9.

A repeated measures analysis of variance was constructed to test the effects of age, fricative consonant, and vowel 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/, /ç/, and /ʃ/), child's age (between subjects; four levels: 2, 3, 4, and 5 years), and vowel context (within subject; five levels: /a/, /i/, /u/, /ɛ/, and /o/), as well as the interaction terms between the three independent variables. Main effects of fricative category, $F(2, 184) = 38.99$, $p < .001$, $\eta_p^2 = .14$, and age, $F(3, 91) = 66.54$, $p < .001$, $\eta_p^2 = .40$, were found. There was a significant Age $\times$ Fricative Category interaction, $F(6, 184) = 4.33$, $p < .001$, $\eta_p^2 = .04$, 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. Although /ç/ and then /ʃ/ are more accurate than /s/ at ages 2 and 3 years, such an advantage is lost at ages 4 and 5 years when /s/ is produced comparably accurately.

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

Age (years)	Vowel	/s/		/ʃ/		/ç/	
		Total	Number (proportion) correct	Total	Number (proportion) correct	Total	Number (proportion) correct
2 (n = 1,139)	/a/	96	21 (.22)	94	43 (.46)	95	70 (.74)
	/ɛ/	0		0		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	
	Overall	380	71 (.19)	377	173 (.46)	382	246 (.64)
3 (n = 1,199)	/a/	100	50 (.50)	100	73 (.73)	100	91 (.91)
	/ɛ/	0		0		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	
	Overall	400	200 (.50)	400	296 (.74)	399	367 (.92)
4 (n = 1,090)	/a/	93	81 (.87)	92	79 (.86)	91	88 (.97)
	/ɛ/	0		0		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	
	Overall	362	297 (.82)	365	326 (.89)	363	356 (.98)
5 (n = 1,197)	/a/	100	90 (.90)	100	93 (.93)	99	98 (.99)
	/ɛ/	0		0		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	
	Overall	398	358 (.90)	400	376 (.94)	399	393 (.98)
Total		1,540	926 (.60)	1,542	1,171 (.76)	1,543	1,362 (.88)

Figure 1. Transcribed accuracy of the three fricatives produced by children as a function of their age. The sound /s/ is represented by red squares, /ʃ/ by blue circles, and /ç/ by green triangles.

No significant overall effect of vowel was found, $F(4, 372) = 2.11$, $p = .08$. No Fricative Category $\times$ Vowel interaction was found, $p = .18$, nor was a Fricative Category $\times$ Vowel $\times$ Age interaction, $p = .48$, suggesting the minimal role vowel context played in the acquisition of these fricatives.

As shown in Table 2, the overall mean accuracy rates were .60 for /s/ (926 correct tokens out of 1,540), .76 for /ʃ/ (1,171 correct tokens out of 1,542), and .88 for /ç/ (1,362 correct tokens out of 1,543). In total, 3,459 tokens out of 4,625 were judged to be correct. The remaining 1,166 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 /ʃ/ or /ç/. In a similar vein, /ç/ was frequently substituted with /s/ or /ʃ/. In addition, despite the early emergence of /ç/, it is not the primary sound that children used to substitute for /s/ or /ʃ/, whereas mispronunciations for /s/ and /ʃ/ are primary substitutions for one another. This suggests that the muscles used for /ç/ versus /s/ and /ʃ/, or the motor-control program for /ç/ versus /s/ and /ʃ/, 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 categories. Despite the high overall interrater reliability (92%), variations exist for the percentage of agreement for different transcribed categories, with the lowest being the /ç/-to-[tʃ^h] 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.

Table 4 lists the age breakdown of the distribution of the top three error types of each fricative target. It illustrates that the percentage of errors declines as children age, except for the case of /ʃ/-to-[s] substitution, where the highest percentage of errors was found at age 3 years 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-acquired sound and is not frequently used to substitute for the other two sounds by younger children.

Table 3. Number (percentage) of the transcribed sounds for the tokens of /s/, /ʃ/, and /ç/ that were judged to be incorrect.

Target fricative	Fricative place		Stopping		Affrication		Deletion
	Error	Number (percentage)	Error	Number (percentage)	Error	Number (percentage)	
/s/ (n = 614)	[ʃ]	331 (54%) [81%]	[t ^h]	9 (1%)	[tʃ]	17 (3%)	17 (3%)
	[ç]	86 (14%) [84%]	[p ^h]	7 (1%)	[tʃ ^h]	14 (2%)	
	[θ]	55 (9%)	[k]	6 (1%)	[ts ^h]	9 (1%)	
	[f]	6 (1%)	[t]	5 (1%)	[ts]	7 (1%)	
/ʃ/ (n = 371)	[s]	104 (28%) [84%]	[t ^h]	9 (2%)	[tʃ]	45 (12%)	22 (6%)
	[ç]	74 (20%) [70%]	[t]	2 (1%)	[tʃ ^h]	17 (5%)	
	[h]	43 (11%) [67%]			[tʃ ^h]	4 (1%)	
	[f]	14 (4%)			[ts]	3 (1%)	
	[θ]	7 (2%)					
/ç/ (n = 181)	[s]	72 (40%) [62%]	[t]	5 (3%)	[tʃ ^h]	29 (16%) [57%]	9 (5%)
	[ʃ]	14 (8%)	[t ^h]	4 (2%)	[tʃ]	11 (6%)	
	[h]	11 (6%)	[k ^h]	2 (1%)	[tʃ]	7 (4%)	
	[θ]	4 (2%)					

Note. Only error patterns that represent at least 1% of the total errors for each target are reported. Brackets indicate percentage of agreement between the two transcribers, when applicable.

Table 4. 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 type	Total number of errors	Age group			
			2 years	3 years	4 years	5 years
/s/	[ʃ]	331	147	141	37	6
	[ç]	86	68	18	0	0
	[θ]	55	37	18	0	0
/ʃ/	[s]	104	19	40	26	19
	[ç]	74	59	11	0	4
	[tʃ]	45	30	15	0	0
/ç/	[ʃ]	72	60	7	0	5
	[tʃʰ]	29	20	9	0	0
	[s]	14	5	5	4	0

Acoustic Analysis

The purposes of the acoustic analysis are (a) to provide an independent assessment of the chronology of fricative acquisition to compare with the transcription results and (b) 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 on the basis of these analyses assumes a definition of acquisition to mean the greatest 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/, /ʃ/, and /ç/), with /s/ being the baseline. The independent variables are normalized values of centroid frequency and F2 onset 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 a 1-unit increase in centroid frequency, the log odds for a fricative to be classified as /ʃ/ (relative to /s/) decrease by 6.889, and the log odds for a fricative to be classified as /ç/ (relative to /s/) decrease by 3.934. Because the baseline is set to be /s/, such decreases are in the predicted direction, because /s/ has a higher centroid frequency than /ʃ/ or /ç/. In a similar fashion, a 1-unit increment in F2 onset frequency enhances the log odds of /ʃ/ by 2.246 and the log odds of /ç/ by 4.111. Again, the higher increase in the probability for /ç/ is expected, because high F2 onset frequency is characteristic of the /ç/ sound. It is also interesting to note that strength of prediction for each acoustic parameter in contributing to different contrasts. For the log odds of /ʃ/ versus /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), whereas the F2 onset is more important for differentiating /ç/ from /ʃ/ (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 /ʃ/, and 90% of /ç/, demonstrating the robustness of the two acoustic parameters in describing Putonghua fricatives. Intrarater reliability was calculated after the first author transcribed 20% of the original data blind. This yielded a score of 98%, which demonstrated the validity of the first-pass transcription.

Children

A similar multinomial analysis was conducted with child speakers. This analysis aims to quantify and predict the order of fricative acquisition by combining the two acoustic parameters. To be specific, 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 5-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 5-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 /ʃ/ and /ç/ over /s/, whereas an increase in F2 onset enhances the likelihood of /ʃ/ and /ç/ relative to /s/. Furthermore, centroid frequency carries greater significance for the /ʃ/-/s/ contrast (2.992 vs. 0.994), and F2 onset is crucial for distinguishing the /ç/-/s/ contrast (3.616 vs. 1.435). Again, both parameters are statistically significant. The similar results between the adults' and the 5-year-olds' models are noteworthy, as they demonstrate that 5-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 on the basis of the 5-year-olds' model.

Table 6a displays the predicted production accuracy in all four age groups on the basis of the statistical model of 5-year-olds' 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 /ç/ is more robustly classified on the basis of acoustic measures than the other two sounds in all four age groups. The sound /ʃ/ ranks second, and the /s/ sound is last.

Although the model has the advantage of providing an independent assessment of the transcription method, it assumes similar vocal-tract lengths between 5-year-old children and younger children, or minimum impact of any differences in vocal-tract length across age groups on the consequent acoustic output. Neither of these assumptions, unfortunately, 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 et al., 2009; Vorperian & Kent, 2007). To avoid making these assumptions, another model

Table 5. Coefficients of the multinomial logistic-regression model conducted for (a) adults and (b) children.

(a) Adults					
Outcome	Variable	Coefficient	Standard error	z value	p value
/ʒ/ vs. /s/	Intercept	0.496	0.338	1.471	<.001
	Centroid frequency	-6.889	0.504	-13.661	<.001
	F2 onset frequency	2.246	0.363	6.178	<.001
/ç/ vs. /s/	Intercept	1.740	0.296	5.873	<.001
	Centroid frequency	-3.934	0.432	-9.091	<.001
	F2 onset frequency	4.1097	0.357	11.491	<.001
Residual deviance: 626.696; Akaike information criterion: 638.696					
(b) Children					
Outcome	Variable	Coefficient	Standard error	z value	p value
/ʒ/ vs. /s/	Intercept	0.152	0.148	1.028	<.001
	Centroid frequency	-2.992	0.193	-15.540	<.001
	F2 onset frequency	0.994	0.161	6.179	<.001
/ç/ vs. /s/	Intercept	-0.051	0.158	-0.318	<.001
	Centroid frequency	-1.435	0.184	-7.799	<.001
	F2 onset frequency	3.616	0.219	16.441	<.001
Residual deviance: 1,045.497; Akaike information criterion: 1,057.497					

Note. In each model, the log odds of the outcomes (three levels: /s/ vs. /ʒ/ vs. /ç/) are modeled as a linear combination of the predictor variables. The dependent variables are the three target fricatives (/s/ as the baseline), and the independent variables are standardized values of centroid frequency and second-formant (F2) onset frequency.

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

Table 6. Percentage of predicted accuracy for each fricative produced by children in each age group, on the basis of two multinomial models.

(a)			
Age group	/s/	/ʒ/	/ç/
2 years	12%	40%	80%
3 years	34%	67%	94%
4 years	73%	70%	95%
5 years	83%	82%	88%
(b)			
Age group	/s/	/ʒ/	/ç/
2 years	11%	53%	70%
3 years	42%	73%	85%
4 years	73%	78%	89%
5 years	84%	89%	88%

Note. In (a), predictions were made by fitting all children's data over the model constructed over 5-year-olds' productions. The dependent variables are the intended/canonical fricative categories, and the independent variables are centroid frequency and F2 onset frequency of the 5-year-olds' productions. In (b), predictions were made on the basis of the model constructed on the data from children in all four age groups. The dependent variables are the transcribed fricative categories, and the independent variables are the centroid frequency and F2 onset frequency of all children's productions. Children's age was coded as a covariate in (b).

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 /ç/ → /ʒ/ → /s/, particularly for the younger age groups. Therefore, although both models are limited in their own way (the 5-year-olds' model assumes equal vocal-tract length, and 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 on the basis of 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 also have contributed to the higher accuracy of the transcription results.

The pattern of emergence of the three fricatives in the two-dimensional acoustic space can be seen in Figure 2 (the relevant summary statistics are in Table 7). At age 2 years, 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 /ʒ/ and /ç/, because they have the acoustic values appropriate for /ʒ/ in the centroid dimension and for /ç/ in the F2 dimension. By age 3 years, /ç/ starts to move to its expected acoustic region surrounding 8000 Hz in centroid frequency and around 3200 Hz for F2 onset.

Figure 2. Scatter plot of fricative productions by adults and children in each age group, plotted according to centroid and F2 onset frequency. The sounds /s/, /ʃ/, and /ç/ are represented, respectively, by red squares, blue circles, and green triangles.

Overlaps remain between /ʃ/ and /s/ until age 4 years, when the two sounds occupy different acoustic areas, primarily differentiated from each other in the centroid dimension (<8000 Hz for /ʃ/ and >8000 Hz for /s/). Comparing against adults, it is also evident that 5-year-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 5-year-olds' instead of adults' speech for statistical modeling.

Children's speech was further depicted for each acoustic dimension to examine the development of different aspects 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 /ʃ/ and the other two sounds around 40 months. A second separation occurs at about 50 months between /s/ and /ç/. In the F2 onset dimension, a distinction between /ç/ and the other two sounds is already present at around 24 months, whereas /s/ and /ʃ/ do not further diverge from each other until after 50 months. The early distinction between /ç/ and /s, ʃ/ 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 /ʃ/ in the centroid dimension.

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 /ç/ → /ʃ/ → /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. The relatively early emergence of /ç/ clearly cannot be accounted for by Jakobson's phonological typology, because this sound occurs infrequently among the world's languages. By contrast, the early acquisition of /ç/ can be more readily attributed to the early maturation of control over the muscles that elevate the tongue body

Table 7. Mean (standard deviation) frequency of the two acoustic parameters, by age group.

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

Figure 3. Scatter plot of averaged centroid and F2 onset values for each fricative token produced by children as a function of their age (in months). Each child is represented by one symbol. Straight lines are the best-fit lines for each fricative, with the dependent variable being the acoustic value and the independent variable being the child's age. The curved lines represent the 95% confidence interval of the best-fit line. The sounds /s/, /ʃ/, and /ç/ are represented, respectively, by red squares, blue pluses, and green triangles.

in producing this sound. However, the oromotor maturational account by itself is unable to explain why the acquisition of /s/ lags behind /ʃ/, if both involve the same articulatory muscle, the tongue tip. The sequence /ʃ/ → /s/ can be better predicted by language-specific phoneme-distribution patterns, because the sound /ʃ/ 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 oromotor 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. To be specific, 2-year-olds do not distinguish among the three fricatives in an adult-like manner. The sound /ç/ first separates from this undifferentiated acoustic space by age 3 years. At age 4 years, both /ʃ/ and /s/ occupy the acoustic range typical for /ʃ/, and the sound /s/ finally separates from /ʃ/ at age 5 years. It is important to note that the initial form of fricatives in 2-year-olds' speech is ambiguous between /ç/ and /ʃ/: It has the acoustic characteristics appropriate for /ʃ/ in the centroid dimension and appropriate for /ç/ 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 interrater agreement for the transcribed sounds (Table 3). Furthermore, as F. 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 interlistener 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 in comparing the current study with that of F. 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 produced 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 adultlike 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 /ç/ in Putonghua-speaking children adds to the small but diverse corpus of literature documenting the consistently early emergence of /ç/ in relation to other sibilant fricatives across languages (Ingram, 1988b; Nakanishi et al., 1972; Zharkova, 2005). 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, in that 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/, whereas Japanese listeners exhibit the opposite pattern (Li, Munson, Edwards, Yoneyama, & Hall, 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 of F2 onset has been shown to enjoy a perceptual advantage in English-speaking children's fricative-perception development over fricative-internal cues (Nitttrouer, 1992, 2002; Nitttrouer & 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, whereas older children and adults pay greater attention to the fricative-internal cue (i.e., centroid frequency). If such a perceptual bias toward F2 onset is universally valid, it provides an alternative account to the oromotor-maturation hypothesis for the early acquisition of /ç/ 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 on the basis of 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.

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, on the basis of our study, we would predict that errors in /ç/ would be less among Putonghua children than would errors in /s/ and /ʃ/, because /ç/ 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 among all three of these fricatives. As a final matter, if our interpretation of the acoustic analysis is correct, then we predict that the accurate production of /ç/ would be related to the ability

to maneuver the tongue dorsum in the coronal plane of the oral cavity. Although such a prediction awaits verification from future direct-imaging studies, results of our study suggest the potential use of /ç/ as a diagnostic marker for speech pathology and point to the possible physiological basis of any potential delay associated with /ç/ production.

Acknowledgments

Data collection and analysis were supported by the Ohio State University Target Investment Fellowship to Fangfang Li and Eunjong Kong, National Institute on Deafness and Other Communication Disorders Grant 02932 to Jan Edwards, and the University of Lethbridge Start-up Fund to Fangfang Li. We thank the staff in Songyuan No. 2 Daycare Center for 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 used 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(2), 18–28.
- Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer* (Version 4.3.07) [Computer software]. Retrieved from <http://www.praat.org/>
- Butterworth, G., Verweij, E., & Hopkins, B. (1997). The development of prehension in infants: Halverson revisited. *British Journal of Developmental Psychology*, 15, 223–236.
- Chew, J. J., Jr. (1969). The structure of Japanese baby talk. *The Journal-Newsletter of the Association of Teachers of Japanese*, 6(1), 4–17.
- de Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16, 1–17.
- Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics and Phonetics*, 22, 937–956.
- Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 51, 419–439.
- Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, 84, 115–123.
- Gibbon, F. E. (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. Fujisaki (Eds.), *Speech production and language: In honor of Osamu Fujimura* (pp. 177–193). Berlin, Germany: Mouton de Gruyter.
- Hosmer, D. W., & Lemeshow, S. (2000). Multiple logistic regression. In D. W. Hosmer & S. Lemeshow (Eds.), *Applied logistic regression* (2nd ed., pp. 31–46). New York, NY: Wiley.

- Hu, F. (2008). The three sibilants in Standard Chinese. In R. Sock, S. Fuchs, & Y. Laprie (Eds.), *Proceedings of the 8th International Seminar on Speech Production* (pp. 105–108). Rocquencourt, France: INRIA.
- Huang, S., Bian, X., Wu, G., & McLemore, C. (1997). *LDC Mandarin Lexicon*. Philadelphia, PA: 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.
- Irwin, O. C. (1933). Proximodistal differentiation of limbs in young organisms. *Psychological Review*, 40, 467–477.
- Jakobson, R. (1968). *Child language, aphasia, and phonological universals* (Keiler, A. R., Trans.). The Hague, the Netherlands: Mouton. (Original work published 1941)
- Jeng, H.-H. (1979). The acquisition of Chinese phonology in relation to Jakobson's laws of irreversible solidarity. In E. Fischer-Jørgensen, J. Rischel, & N. Thorsen (Eds.), *Proceedings of the 9th International Congress of Phonetic Sciences: Vol. 2* (pp. 155–161). Copenhagen, Denmark: University of Copenhagen.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108, 1252–1263.
- Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65–89). Timonium, MD: York Press.
- Kier, W. M., & Smith, K. K. (1985). Tongue, tentacles and trunks: The biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society*, 83, 307–324.
- Ladd, D. R. (2011). Phonetics in phonology. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The handbook of phonological theory* (2nd ed., pp. 348–373). New York, NY: Blackwell.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, United Kingdom: Blackwell.
- Ladefoged, P., & Wu, Z. (1984). Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics*, 12, 267–278.
- Lee, W.-S. (1999). An articulatory and acoustical analysis of the syllable-initial sibilants and approximants in Beijing Mandarin. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 413–416). College Park, MD: American Institute of Physics.
- Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4, 185–199.
- Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese, and Mandarin Chinese* (Unpublished doctoral thesis). Ohio State University, Columbus, OH.
- Li, F. (2012). Language-specific developmental differences in speech production: A cross-language acoustic study. *Child Development*, 83, 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, 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. *The Journal of the Acoustical Society of America*, 129, 999–1011.
- 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.
- Lieberman, 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 phonetics]. Beijing, China: Peking University Press.
- Lin, Y. S., & Peng, S. C. (2003). Acquisition profiles of syllable-initial consonants in Mandarin-speaking children with cochlear implants. *Acta Oto-Laryngologica*, 123, 1046–1053.
- Locke, J. L. (1983). *Phonological acquisition and change*. New York, NY: Academic Press.
- Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7, 41–74.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge, United Kingdom: 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, 147–154.
- McEnery, A., & Xiao, Z. (2004). The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Fourth International Conference on Language Resources and Evaluation* (pp. 1175–1178). Paris, France: ELRA–European Language Resources Association.
- McGowan, R. S., & Nittrouer, S. (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/. *The Journal of the Acoustical Society of America*, 83, 229–236.
- 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 & Phonetics*, 24, 245–260.
- Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, 21, 124–139.
- Nakanishi, Y., Owada, K., & Fujita, N. (1972). Kōon kensa to sono kekka ni kansuru kōsatu [Translated title]. *Tokushū kyōiku kenkyū shisetu hōkoku* [Bulletin of the Tokyo Gakugei University Special Education Research Group], 1, 1–41.
- 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. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *The Journal of the Acoustical Society of America*, 97, 520–530.
- Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America*, 112, 711–719.
- Nittrouer, S., & Miller, M. E. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *The Journal of the Acoustical Society of America*, 102, 572–580.
- Pye, C., Ingram, D., & List, H. (1987). A comparison of initial consonant acquisition in English and Quiché. In K. E. Nelson & A. van Kleeck (Eds.), *Children's language: Vol. 6* (pp. 175–190). Hillsdale, NJ: Erlbaum.

- R Core Team.** (2011). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/>
- Rahim, K. J., Burr, W. S., & Thomson, D. J.** (2014). *Applications of multitaper spectral analysis to nonstationary data* (Doctoral dissertation). Retrieved from QSpace at Queen's University (<http://hdl.handle.net/1974/12584>).
- Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P.** (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 194–207). Cambridge, United Kingdom: Cambridge University Press.
- Shadle, C. H., & Mair, S. J.** (1996). Quantifying spectral characteristics of fricatives. *Proceedings of the International Conference on Spoken Language Processing (ICSLP96)*, 1517–1520.
- Shih, Y.-T., & Kong, E.** (2011). Perception of Mandarin fricatives by native speakers of Taiwan Mandarin and Taiwanese. In Z. Jing-Schmidt (Ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics: Vol. 1* (pp. 110–119). Eugene, OR: University of Oregon.
- 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.
- Stevens, K. N., Li, Z., Lee, C.-Y., & Keyser, J.** (2004). A note on Mandarin fricatives and enhancement. In G. Fant, H. Fujisaki, J. Cao, & Y. Xu (Eds.), *From traditional phonology to modern speech processing* (pp. 393–403). Beijing, China: Foreign Language Teaching and Research Press.
- Stone, M.** (1990). A three-dimensional model of tongue movement based on ultrasound and X-ray microbeam data. *The Journal of the Acoustical Society of America*, 87, 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 and Hearing Research*, 28, 455–462.
- Svantesson, J.-O.** (1986). Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics*, 14, 53–70.
- Toda, M., & Honda, K.** (2003, December). *An MRI-based cross-linguistic study of sibilant fricatives*. Paper presented at the 6th International Seminar on Speech Production, Sydney, Australia.
- Tsoi, W. C. T.** (2005). *The effects of occurrence frequency of phonemes on second language acquisition: A quantitative comparison of Cantonese, Mandarin, Italian, German, and American English* (Master's thesis). Chinese University of Hong Kong.
- 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, 27–48.
- 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.
- Vorperian, H. K., & Kent, R. 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., ... Gentry, L. R.** (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*, 125, 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, 1912–1918.
- 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 - 一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. *Leiden Working Papers in Linguistics*, 2(1), 189–213.
- Zhu, H., & Dodd, B.** (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27, 3–42.

Appendix

The Word List

Target vowel	/s/	Chinese orthography	/ʈʂ/	Chinese orthography	/ʃ/	Chinese orthography
/a/	/sa.niao/	撒尿	/ʈʂa.jy/	下雨	/ʃa.fa/	沙发
	/san/	伞	/ʈʂaŋ.tsi/	箱子	/ʃan/	山
	/san.tɕiao/	三角	/ʈʂaŋ.tɕao/	香蕉	/ʃan.yaŋ/	山羊
	/sa.lə/	撒了	/ʈʂa/	虾	/ʃan.tsi/	扇子
/i/	/si.tɕi/	司机	/ʈʂi.ʂo/	洗手	/ʃi.tsi/	石子
	/si.kə/	四个	/ʈʂi/	细	/ʃi/	十
	/si/	丝	/ʈʂi.kua/	西瓜	/ʃi.tsi/	狮子
	/si.miao/	寺庙	/ʈʂiŋ.ʈʂiŋ/	星星	/ʃi.tsi/	柿子
/o/	/soŋ.su/	松鼠	/ʈʂo.ʈʂi/	休息	/ʃo.tɕyan/	手绢
	/soŋ/	松	/ʈʂo/	袖	/ʃo.tʰao/	手套
	/soŋ.tɕin.tai/	松紧带	/ʈʂoŋ.mao/	熊猫	/ʃo.tsi/	瘦子
	/soŋ.lə/	松了	/ʈʂoŋ/	熊	/ʃo/	手
/u/	/sun/	笋			/ʃu.pao/	书包
	/su.liao/	塑料			/ʃu.su/	叔叔
	/sun.tsi/	孙子			/ʃu/	树
	/sun.wu.koŋ/	孙悟空			/ʃu.tsʰai/	蔬菜
/ɛ/			/ʈʂɛn.hua/	鲜花		
			/ʈʂɛn/	线		
			/ʈʂɛ.tsi/	写字		
			/ʈʂɛ.tsi/	鞋子		