THE PHONETIC DEVELOPMENT OF VOICELESS SIBILANT FRICATIVES
IN ENGLISH, JAPANESE AND MANDARIN CHINESE

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Fangfang Li, B.A.

*****

The Ohio State University
2008

Dissertation Committee: Approved by

Mary E. Beckman, Advisor

Susan Nittrouer   _____

Cynthia Clopper   Advisor

                                    Graduate Program in Linguistics

ABSTRACT

This dissertation examines the development of voiceless sibilant fricatives in children speaking English, Japanese or Mandarin Chinese. Both English and Japanese have a two-way distinction in sibilant fricatives (/s/ vs. /ʃ/ in English and /s/ vs. /ɕ/ in Japanese), and Mandarin Chinese has a three-way contrast among /s/, /ɕ/ and /ʂ/. Children's fricative productions have been traditionally described using adult's impressionistic transcriptions, which yield inconsistent orders of acquisition both across children and across languages. This dissertation argues that transcription filters children's early productions through adults' language-specific phonological systems, and therefore obscures the actual developmental patterns in children's speech. The purpose of the current study is to tease apart children's own productions from adults' interpretations of them by applying acoustic analyses to both adults' and children's productions and by systematically evaluating adults' perception patterns in tasks that allow more continuous responses. This dissertation first starts by examining the acoustics of adult productions in all three languages to parameterize the acoustic space for sibilant fricatives. It then investigates the production patterns of children speaking either English or Japanese, which has a two-way contrast in voiceless sibilant fricatives. Twenty 2-to-3-year old speakers of each of the two languages were tested to look for covert contrast in children's speech. The results show that adults are not able to recognize fine-grained phonetic differences that children make, and a more objective description of children's productions using methods such as acoustic

analysis is needed. A set of perception experiments was then performed to further examine how English-speaking adults and Japanese-speaking adults would differ in judging these 2-3 year olds' productions. The results of this set of perception experiments show that English adults and Japanese adults correlate different acoustic cues with their fricative categories. These production and perception experiments were further extended to other age groups and to Mandarin Chinese in order to make more robust generalizations on the crosslinguistic production and perception patterns. The results suggest that children's early productions are intermediate and variable, with no clear category distinctions, and adults categorize these gross productions in language-specific ways. That is, children start by making some undifferentiated lingual gestures in the multidimensional acoustic space, and as their age increases, they separate out categories in the parameters that are salient in adult productions, which are also cues that adults use in judging children's productions.

ACKNOWLEDGMENTS

<div align="center">

VITA

</div>

2001 .... B.A., English literature and linguistics,
Beijing University, Beijing, China

2001 - 2003 .... Teaching Assistant,
The Ohio State University

2003 - 2007 .... Research Assistant,
The Ohio State University

<div align="center">

PUBLICATIONS

</div>

1. Li, F., Edwards, J., & Beckman, M. E. (in press, 2008). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. Journal of Phonetics.

<div align="center">

FIELDS OF STUDY

</div>

Major Field: Linguistics

Specialization: Phonetics and child speech acquisition

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure   **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1   Child speech development and phonological universals

Speech, as one of the unique aspects of human beings, distinguishes us from the other animals in the world (Hockett, 1960). The ability to communicate using speech, however, does not come without effort – infants are not born ready to communicate with adults. In fact, both parents and newborns are easily frustrated by the lack of a communicative channel (Irwin and Briggs-Gowan, 2002). Although speech may seem to be automatic and effortless for adults, they very often forget how long and how much effort they have spent in order to use it at will. Compared with how much we know about the world today (Darwin, 1859; Einstein, 1916), our knowledge is astonishingly limited in regard to how we ourselves develop from infants to adults while acquiring language at the same time in order to become part of a specific language-speaking community. The study of child speech development can not only inform us of the ontology of language in individual, shedding light on historical sound changes, but also unveil the developmental trend that may underly all common cognitive processes.

The early records of children's phonological acquisition are from diary studies that describe children's speech in chronological order (Darwin, 1887; Heroard, 1868). It is found that children are not initially able to produce all the sounds in their native language. Nor is it completely random which of these sounds they produce first.

Rather, they bootstrap from a few "simple" sounds, and use them to substitute for the more "difficult" ones, before they master all the sounds in their native inventory. Also the order in which children acquire different sound classes crosslinguistically seems to be largely comparable. More specifically, most children can produce the vowels in the ambient language by about age 2. Also they develop stop consonants and glides relatively early, while fricatives, affricates, and liquids tend to be later-acquired (Locke, 1983). In his pioneering book, Jakobson attributed the similarities observed across languages to universal substantive principles - "implicational laws" - that structure the phoneme inventories of all spoken languages and that also determine how children acquire speech sounds (Jakobson, 1941/1968).

Although the relative chronology of the emergence of sound classes has been shown to be largely invariant across languages, the question remains as to whether universal acquisition sequence exists inside each sound class. Jakobson would argue the affirmative by stating in his 1941/1968 book: "At a particular stage of development, ..., the Swedish child says *tata* for 'kaka', the German child *topf* for 'kopf', the English child *tut* for 'cut', and the Japanese child also changes *k* to *t*." This is a strong claim that not only assumes sounds *k* or *t* are similar across languages but also proposes that there is an inherent acquisition order between these specific sounds that is also universal. Moreover, Locke (1983), in his cross language survey on child phonological development, claims fronting is a cross-linguistic developmental pattern inside the fricative class, with front fricatives being usually substituted for the back fricatives.

It has to be mentioned that this early hypothesis of the universal acquisition order being fully specified inside each sound class is mostly based on descriptions of early diary studies. It did not get support from longitudinal studies over a larger sample of children or from large-scale cross-sectional norming studies that sampled a

large population of children in different age groups in order to establish acquisition norms of typical-developing children (Wellman, Case, Mengert, and Bradbury, 1931; Poole, 1934; Prather, Hedrick, and Kern, 1975). The major method used in these studies is similar to what people initially used in the diary studies – to label the sounds in terms of adult phonological categories. These studies show that while the fronting from k to t is a common pattern found in English-speaking children, it is not the case for Japanese-speaking children, who are more likely to substitute [k] for /t/ (Nakanishi, Owada and Fujita 1972). Fricative acquisition poses a more serious question to this claim of universal acquisition order, as they are generally late acquired by children and the acquisition order in fricatives is controversial across languages, and sometimes within a certain language.

For example, for English, large norming studies report that the typical error pattern for voiceless sibilant fricatives is to front /ʃ/ to /s/ (Poole, 1934; Templin, 1957; Smit, 1990). Other studies such as Ingram (1981), however, found that /s/ was acquired later than /ʃ/, for example, if a narrower transcription is used such that only adult-like /s/ productions as identified by native transcribers were counted as correct in the transcription. Moreover, Prather et al. (1975) noticed a reversal phenomenon in the production of the /s/ sound, which was reported to reach the designated percentage correct standard in a earlier age, but dropped below that standard at a later age level. In Japanese, according to Nakanishi et al. (1972), /s/ is among the most error-prone sounds for 4-6 year olds, and it is a common error to substitute [ɕ] or [tɕ] for /s/. Similarly, Yasuda (1966) shows that /ɕ/ is acquired earlier than /s/. Nishimura (1980) did a longitudinal study on four Japanese-speaking children in spontaneous play settings, and found that /s/ and /ɕ/ seem to emerge in children of similar ages. For Mandarin, the literature disagrees on the relative order of acquisition for the three sibilant fricatives. For example, Zhu and Dodd (2000) did a normative

cross-sectional study on 129 Mandarin-speaking children aged 1;6 to 4;6, and found that /ɕ/ occurs the earliest among the three sounds and there is a tendency for /s/ to emerge earlier than /ʂ/. Li, Zhu, Dodd, Windsor, Kelly, and Hewlett (2002), however, examined 4 children longitudinally, and found that /ɕ/ emerged first for two children, and /s/ for the other two.

There are two possible reasons of why these large scale studies fail to show any universally consistent order of acquisition inside the sibilant fricative class. One possibility is simply that such universal order does not exist and there is no universality in how children acquire speech sounds inside each sound class. The other possibility is that there still exists some universal trend in children's speech development, which is obscured by the traditional transcription method used to describe children's speech. This dissertation aims to test the second possibility by using acoustic methods to describe productions by a large number of children in different languages.

There are three reasons why the traditional method – transcription or labeling – is likely to give rise to inconsistent and unreliable results in describing children's production patterns both within a certain language and across languages. For one thing, adults do not always agree with each other especially over unclear speech. Unfortunately, children's early speech productions are very often the case of such unclear speech. Even the same adult may have different opinions on a single sound heard at different times (Kent, 1996). Also depending on the amount of phonetic training a transcriber has, the dialect and the number of foreign languages s/he speaks, the categorical judgment over ambiguous tokens will yield different results for different people. Thus no matter how well controlled, this method is inevitable to have some degree of subjectivity.

The second reason is that transcription is unsuitable to make crosslinguistic comparisons for children's productions. This is because sounds labeled with the same

symbol may differ in fine-grained articulatory and acoustic details in different languages. As technology advances, many studies have shown that sounds transcribed the same way across languages can differ significantly in terms of their articulatory and acoustic details (Cho and Ladefoged, 1999; Toda and Honda, 2003). For example, many languages have the sound that is transcribed as [r], however, the trilled articulation of [r] in Spanish is by no means identical to that of the [r] in English. Another example is that both English and Japanese have the sound that labeled as /t/, but the VOT (Voice Onset Time) range for what counted as a /t/ in the two languages differ greatly from each other (Riney, Takagi, Ota, and Uchida, 2007; Kong, 2009). Children's productions of similar sounds in different languages may later be categorized into different phonological categories by adult transcribers.

The third reason why the transcription method may blur the true devleopmental pattern is because even in the case where adult transcribers perfectly agree with each other in describing certain sounds, transcription is still an indirect measure of children's productions, as it tries to fit children's raw productions into adults' well-formed rigid categories. While transcription is fast and convenient (live transcription is still widely used in the clinical field), it fails to capture the fine-phonetic details that are well below adults' categorical perceptual threshold, and hence may miss important developmental patterns that are unique to children. That is, there may well be universal patterns evident in the fine-grained details that are obscured by the way that the phonological systems of different languages "warp" the perceptual phonetic space (Best, McRoberts, and Goodell, 2001).

There are at least two pieces of evidences which suggest the existence of such lower-level phonetic details; a closer examination of them may reveal universal patterns. First, a number of developmental psychologists suggest that children's early productions are often characterized by undifferentiated entities (Ferguson, 1986; Menn

5

and Butterworth, 1983; Studdert-Kennedy, 1991). For instance, Kent (1976) reported that English-children's VOT productions exhibit an undefined unimodal distribution, which will gradually become bimodal. Moreover, at the initial shaping stage of this bimodal distribution, most productions will fall into adults' perceived unaspirated category. It is likely that early fricative productions may also be represented as gross gestures involving tight constriction somewhere in the oral cavity, and does not precisely fit into any specific categories that adults already have. Therefore, transcribing these early productions from children will mean forcing adult transcribers to categorize these intermediate productions into their well-formed categories.

Secondly, there is a growing body of literature describing children's ability to produce small differences that are well under adult's categorical perception threshold, in a pattern known as covert contrast. More formally, covert contrast, also called subphonemic contrast, refers to perceptually indistinguishable, but statistically significant acoustic difference in children's productions of contrastive sounds (Macken and Barton, 1980; Maxwell and Weismer, 1982; Scobbie, Gibbon, Hardcastle, and Fletcher, 2000: among others). For example, Macken and Barton (1980) examined four normal-developing children longitudinally in their productions of VOT as a contrastive cue to aspiration in producing voiceless stops of English. They found covert contrasts in all four children's production of stops before their productions can be perceived as two separate categories. Baum and McNutt (1990) examined the temporal, amplitudinal and spectral representations of /s/ and /θ/ in normal developing children aged 5-8 and compared them with the productions of fronted /s/ and /θ/ in children with frontal misarticulations, finding a statistically reliable difference in both amplitudinal and spectral measures between fronted /s/ and target /θ/ in the misarticulating group. The phenomenon of covert contrast calls for the use of a more objective way to describe children's productions that is not biased by the categories

6

of adult language. In this dissertation, acoustic analysis will be used, as it is non-invasive and less expensive relative to articulatory methods. At the same time, it is well-established and has been proven to be effective in describing productions of both adults (Heinz and Stevens, 1961; Hughes and Halle, 1956: among others) and children (Soli, 1981; Nittrouer, Studdert-Kennedy, and McGowan, 1989).

The developmental patterns that these studies suggest are phonetic universals instead of phonological universals. Based on these studies, I hypothesize that children do not start by producing some well-formed prototypical categories in a discrete manner but rather start from some undifferentiated lingual gestures that are not specified and may be intermediate, and then gradually separate out into different distinct categories.

1.2   Objectives

This dissertation will then focus on describing children's speech development using acoustic analysis, as well as evaluating adult's perception, in order to find out the developmental patterns of voiceless sibilant fricative acquisition within different languages, and to compare them across languages. The goals of this dissertation, therefore, are two-fold. The first is to look at the phonetic development of children's voiceless sibilant fricatives in different languages, and the relationship between children's and adults' productions. The second goal is to examine adults' perceptions and how they relate to children's fricative acquisition patterns.

Moreover, two hypotheses are tested: 1) In terms of phonetic development, children will start by producing some undifferentiated gesture in the multi-dimensional parametric space in fricative acquisition, no matter how many contrasts they have in their native languages. And this undifferentiated gesture will gradually separate

into 2 or 3 distinct gestures that correspond to the contrasting fricatives in the ambient language. 2) In categorizing these early immature child productions, adults are constrained by the language-specific perceptual space for sibilant fricatives. These biases introduced during adults' categorization contribute to the disagreement about fricative acquisition ordering among different languages.

The outline of this dissertation is as follows: Chapter 2 introduces the target sounds – voiceless sibilant fricatives – in detail, by reviewing the literature on the articulation and phonological distributions of sibilants in the three languages to facilitate later discussions on acoustic descriptions. Chapter 3 discusses the experimental results on adults' production patterns in Mandarin, English and Japanese. These production norms constitute the ambient language environment that children are exposed to, and will serve as the baseline for crosslinguistic comparisons in children. Chapter 4 and 5 are about children's production pattern and adults' perception pattern in languages that have a two-way contrast, namely, English and Japanese. Chapter 6 and 7 then expand the investigation into three languages by including Mandarin which has a three-way contrast in voiceless sibilants. Chapter 8 summarizes the findings described in the previous chapters and evaluates their support for the ideas about the domain of universality described above.

CHAPTER 2

VOICELESS SIBILANT FRICATIVES IN THE THREE LANGUAGES

2.1   Introduction

This chapter introduces the subject matter of this dissertation – voiceless sibilant fricatives – in the three languages of interest. The three languages tested all have more than one sibilant fricative. English and Japanese both have an anterior fricative contrasting with a more posterior one, and Mandarin Chinese has a three way contrast involving differences in tongue position and tongue posture. In all three languages, the sibilant fricatives are mastered later relative to nasals and stops (Smit, 1990; Nishimura, Ingram, Peng, and Dale, 1980; Zhu and Dodd, 2000), supporting the idea that they are hard to acquire.

   The reason that fricatives are hard for children to acquire is that they require fine-articulatory control of the tongue, with which children are not readily equipped initially, and have to learn through practice (Kent, 1992). This chapter describes these fricatives in terms of articulatory gestures and the articulatory aspects that adult speakers of different languages control to make the language-specific contrasts. In this dissertation, however, no direct articulatory measures were taken since they are costly and often unsuitable for children. The literature review in this chapter on the articulatory descriptions of sibilants then mainly serves as the background information to facilitate later discussion of the acoustic measures used.

2.2 Articulation of sibilants

Sibilant fricatives are produced with the tongue and the roof of the mouth forming a narrow constriction in the oral cavity, which creates turbulence when the rapid air stream passes through. They are also called '"obstacle fricatives" in Ladefoged and Maddieson (1986), since the production of this set of sounds involves the obstruction of the already turbulent flow by the downstream "obstacle" of incisors, generating a high-pitch hissing sound quality, characteristic of sibilant fricatives. Important to note here is that the fricative spectrum primarily reflects the resonance of the front cavity, the cavity in front of the narrowest constriction. Moreover, different tongue shapes or tongue positions may result in different configurations of the vocal tract, which can potentially give rise to acoustically or perceptually distinct categories (Shadle, 1991). In order to facilitate descriptions of the articulatory gestures involved, some terms regarding different parts of the tongue and the palate will be briefly introduced.

Traditionally, consonant articulations are described from a midsagittal view, as if the head is dissected into halves from front to back, and therefore only the central lines of the tongue and of the roof of the mouth are described. In the midsagittal plane, the tongue is divided into tip, blade, and main body. The definitions of the tongue tip and blade follow Ladefoged (1989), which specifies the tip to be the front part of the tongue that is parallel to the back surface of the teeth plus a small area of about 2 mm on the upper surface (pp.47). Articulation involving the tip of the tongue is called "apical". The blade is defined as the upper surface part of the tongue that extends 8-10 mm behind the tip, with the associating gestures called "laminal". Sounds produced with the tip or the blade are generally referred to as "coronal" in many phonetic and phonological studies (Keating, Paradis, and Prunet, 1991). The main body of the tongue is called the dorsum, following Catford (1988), with

articulations involving the front part of the body being called anterodorsal and the back called posterodorsal. Also, sometimes the tip of the tongue is raised to create a space between the tip and the lower floor of the mouth, which is termed as the "sublingual cavity" (Johnson, 1997; Hamann, 2003).

The productions of sibilant fricatives are commonly described in terms of the "place of articulation", which refers to the location of the narrowest constriction made by the tongue toward the ceiling of the vocal tract in the mid-sagittal plane. For example, "dentals" refer to productions involving some degree of approximation toward the upper incisors. The concave-shaped bony structure right behind the upper teeth before the palate starts is called the "alveolar ridge". Articulations made toward the alveolar ridge are called "alveolar". Those falling into the region between the posterior range of the alveolar ridge and the front of the soft palate are called "post-alveolar". Across languages, /s/ is described as a dental/alveolar fricative. The post-alveolar area contains fricatives which vary so much that further classifications are needed to differentiate them. These categories include palato-alveolar, alveolopalatals, and retroflexes. It has been noted that these sub-categories that make up the post-alveolar class do not represent finer discrete location distinctions, but rather differ from each other in aspects such as tongue posture (Ladefoged and Wu, 1984; Keating et al., 1991). According to Keating et al. (1991), palato-alveolar constrictions are at or around the alveolar ridge, as represented by English /ʃ/. Alveolopalatal fricatives involve the tongue blade raising toward the alveolar ridge and are most often accompanied by bunching of the tongue body as well. The voiceless alveolopalatal fricative is usually represented by /ɕ/, as in Polish and Mandarin Chinese (Ladefoged and Wu, 1984; Lindblad and Lundqvist, 1995; Halle and Stevens, 1997; Stevens, Li, Lee, and Keyser, 2004). As to retroflex sibilants, more variations were documented with some involving the tongue tip raising toward the palatal region, as in the Toda language,

whereas others have a flat tongue shape and a more posterior position as in Tamil (Ladefoged and Maddieson, 1986; Hamann, 2003). The Mandarin Chinese retroflex fricative /ʂ/ has a even flatter tongue shape and smaller sublingual cavity than the other types of retroflex sounds in other languages.

English has two voiceless sibilant fricatives: /s/ and /ʃ/, with /s/ being traditionally described as alveolar and /ʃ/ as rounded palato-alveolar (Ladefoged and Maddieson, 1986; Keating et al., 1991; Akamatsu, 1997). Dart (1998) examined productions of /s/ sounds by 20 American English speakers using static linguagrams and palatograms and found considerable variability among speakers, with 42.5% of the tokens produced as apical and 52.5% as laminal. Unfortunately, Dart's stimuli only contain /s/ sounds in word-medial and word-final positions, preceded or followed by just one vowel, /æ/. Nonetheless, apicality or laminality does not seem to be relevant for differentiating the articulatory gestures for the two sounds of English. The contrast is more commonly believed to lie in the different place where the major lingual constriction is located in the oral cavity, with the constriction for /ʃ/ being further back than that for /s/ (Ladefoged, 1957; Catford, 1988).

In addition, it is noted that both /s/ and /ʃ/ in English are grooved, which means that they both involve a mid-sagittal groove along the central line of the tongue body (Halle, Stevens, Lindblom, and Ohman, 1979). Catford (1988), however, points out that it is in effect the tongue shape in the cross-sectional plane, the cross-sectional area of the tongue, that differentiates different voiceless sibilant fricatives. More recently, Stone, Faber, Raphael, and Shawker (1992) used ultrasound technology to scan the cross-sections of the tongue at different locations, from anterior to posterior, and confirmed that /s/ is produced with a midsagittal groove along the entire length of the tongue, whereas the groove for /ʃ/ is mainly in the the posterior part of the tongue. Fletcher and Newman (1991) did a production experiment using electropalatography

and found that the constriction place and the width of the sibilant groove jointly distinguish the /s/ and /ʃ/ productions. In addition, Ladefoged and Maddieson (1986) point out that English /s/ and /ʃ/ also differ from each other in terms of the posture of the articulator, namely the tongue. More specifically, Ladefoged and Maddieson (1986) terms the /ʃ/ sound in English as domed palato-alveolar in that the tongue shape right behind the primary constriction has a domed shape in relation to the upper roof of the mouth as a result of the tongue front raising, as opposed to the convex shape of the tongue in producing /s/. They further equate this domed shape in /ʃ/ with "a small amount of palatalization".

To summarize, previous articulatory literature on English voiceless sibilant fricatives shows that /s/ and /ʃ/ systematically differ in the following 4 aspects: 1) the primary constriction in the oral cavity, where /s/ is made further front than /ʃ/, 2) the cross-sectional area, with the tongue having a narrower width of the groove in producing /s/ than in producing /ʃ/, 3) the posture of the the tongue, with that for /s/ having a convex shape and that for /ʃ/ a semi-palatalized domed shape, and 4) secondary articulation, where /ʃ/ is rounded and /s/ is not. One thing to note is that with the tongue being a soft flexible tissue (Kier and Smith, 1985; Stone et al., 1992), some articulatory aspects of the tongue may not be completely independent from each other. For example, raising the front of the tongue while semi-palatalizing must yield a wider midsagittal groove, as in the production of /ʃ/.

Japanese /s/ is perceptually very similar to the English /s/, and if there is any difference, it is that the English /s/ is more sibilant than the Japanese one (Akamatsu, 1997). It has to be noted that the Japanese post-alveolar fricative /ɕ/ has been described both as /ʃ/ (Funatsu, 1995) and as /ɕ/ (Akamatsu, 1997; Toda, 2005). However, according to Akamatsu (1997), the Japanese postalveolar fricative and the

English postalveolar /ʃ/ are very different. This is not only because the English /ʃ/ involves lip rounding whereas the Japanese /ɕ/ does not, but also because in producing the Japanese /ɕ/, the tongue predorsum is raised toward the palate, and thus forms a long palatal channel, in contrast with the lack of such channel in producing the /ʃ/ sound. As a result, the Japanese /s/ and /ɕ/ primarily contrast in palatalization (Toda and Honda, 2003; Toda, 2005). Toda and Honda (2003) did a cross-linguistic articulatory study on sibilant fricatives, taking mid-sagittal MRI scans on the fricative productions by speakers of Japanese, English, French, Mandarin Chinese, and Swedish. They designed two articulatory measures: the palatalization index and the place index. The palatalization index corresponds to the average distance between the tongue and the palate. The area is delimited by two reference points: tightest point of tongue constriction and the line in the back which uses the posterior nasal spine as reference. The distance is calculated on the midsagittal plane (see Figure 2.1). The place of articulation index is the area of the part of the oral cavity in front of the major lingual constriction. The results shows that, in English and French, /s/ and /ʃ/ mainly differ in the place index, with varying degrees of palatalization that overlap considerably for the two sounds. In Japanese, however, /s/ and /ɕ/ do not differ so much in the place index; it is the palatalization index that clearly differentiates the two. Because of the convincing results from Toda and Honda (2003), I will use the symbol /ɕ/ to represent the Japanese post-alveolar fricative.

Chao (1948, 1968) speculated that the three voiceless sibilant fricatives in Mandarin Chinese are /s/ being a dental sibilant, /ɕ/ being a palatal, and /ʂ/ being an apical retroflex. Lee (1999) did an articulatory study on the three sibilants using palatograms and linguagrams, and showed that /s/ is an apical or laminal dental-alveolar or alveolar fricative, /ɕ/ is a laminal or anterodorsal postalveolar fricative, and /ʂ/ is an apical postalveolar fricative. Ladefoged and Wu (1984) examined these
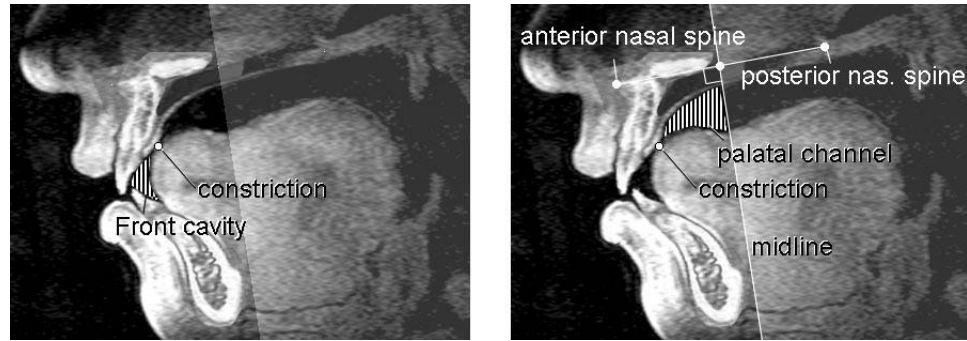
Figure 2.1: Place index (left) and palatalization index (right) in Toda and Honda (2003)

three fricatives using palatograms and midsagittal X-ray photographs, and found that all three of their subjects produced /s/ with the tip of the tongue, and they all show a concave tongue shape in the mid-sagittal plane, with the primary constriction varying from teeth to alveolar ridge. The deep hollowing of the tongue in producing /s/ explains the bigger values of palatal channel area that Toda and Honda (2003) found in Chinese. As for /ʂ/, all three speakers produced it using the upper surface of the tip of the tongue, with the major constriction being at about the center of the alveolar ridge. Also, one notable difference between /s/ and /ʂ/ is that both the height and width of the channel are greater for /ʂ/ than for /s/, which is similar to the English /s-ʃ/ contrast in terms of the cross-sectional area of the anterior groove. As for /ɕ/, the constriction is between that of /s/ and that of /ʂ/, and it differs from both /s/ and /ʃ/ in that it is much higher in the mouth, with the tongue forming a long flat channel with the palate. As a result, the Mandarin sound /ɕ/ differs from the other two sounds in terms of the tongue shape, while /s/ and /ʂ/ contrast in tongue constriction place.

		Postalveolar		
	Dental/alveolar	Palato-alveolar	Alveolo-palatal	Retroflex
English	s (s)	ʃ(S)		
Japanese	s (s)		ɕ(c})	
Mandarin	s (s)		ɕ(c})	ʂ(S)

Table 2.1: Voiceless sibilant fricatives in the three languages described in terms of place of articulation

Table 2.1 summarizes the different places of articulation for the voiceless sibilant fricatives that are present in English, Japanese and Mandarin Chinese. More specifically, both English and Japanese have a two-way voiceless sibilant fricative contrast, with English contrasting /s/ from /ʃ/, and Japanese contrasting /s/ from /ɕ/. Mandarin Chinese has a three-way contrast among /s/, /ɕ/ and /ʂ/. All the fricatives in Table 2.1 are described using IPA symbols with worldBet symbols in parentheses. WorldBet symbols are ASCII encodings of the International Phonetic Alphabet (Hieronymus, 1994), and are used in all the figure captions later in this dissertation.

As pointed out by Ladefoged and Wu (1984) and Keating et al. (1991), the term "place of articulation" for sibilant fricatives is a complex property involving differences in tongue place and tongue posture as well as other "secondary" features such as lip rounding versus spreading. And indeed, the term "retroflex" itself names a posture of the tongue tip rather than any specific post-alveolar place. Thus, different sibilants are produced with different articulatory gestures, despite sharing the coronal feature. Moreover, even those sibilants that are labeled with the same phonetic symbol in different languages may not be identical in articulatory details. These differences in articulatory terms can be related to various acoustic descriptions, as will be shown in Chapter 3.

2.3 Phonological distributions of sibilants

The sibilant fricatives in the three languages not only differ in articulatory configurations, but also in the phonological environments in which they can occur. In English, /s/ and /ʃ/ both occur in front of all vowels, and thus share a completely overlapping distribution. In Japanese, /s/ can never occur before vowel /i/. Alveolopalatal /ɕ/ can occur before all vowels, but its occurrence before /e/ is very marginal, being limited to recent loan words such as /ɕerii/ 'sherry' (Akamatsu, 1997). In Mandarin Chinese, /s/ and /ʂ/ rarely occur in the /e/ context, and neither of them can occur before the vowel /i/, but each occurs before a homorganic "apical vowel" which is transcribed as [ɹ] after /s/ and [ɻ] after /ʂ/. These three vowels – [i], [ɹ] and [ɻ] – are all written with "i" in the Pinyin writing system, suggesting an analysis whereby all are allophones of a high, non-rounded category. However, only [i] after /ɕ/ is a properly "front" vowel. Further, /s/ and /ʂ/ can both occur in front of the vowel /u/, but /ɕ/ can only occur in front of the vowel /y/ instead. As a result, it is only in the context of the vowels /a/ and /o/ that all three sibilant fricatives in Mandarin are robustly contrastive.

2.4 Frequency distributions of sibilants

Phoneme frequency as well as consonant-vowel biphone frequency are other factors that can potentially influence children's fricative acquisition. They may play a role in shaping children's fricative perception and production, and may also affect how adults perceive the ambiguous fricatives that are so characteristic of children's productions. If one fricative is more frequently occurring than the other, children are likely to get more input from the ambient language which may facilitate their category formation.

Also the more frequent fricative is likely to bias adults in perceiving ambiguous speech produced by children.

The frequency information for English and Japanese is taken from Edwards and Beckman (2008), and the calculations are based on adult online lexicons. The corpus used for English is the Hoosier Mental Lexicon (HML, Pisoni, Nusbaum, Luce, & Slowiacek, 1985), which is a list of about 19,000 word form types based on Webster's Pocket Dictionary. For Japanese, it is a subset of words from the NTT database (Amano & Kondo, 1999), which is based on the third edition of the Sanseido Shinmeikai Dictionary (Kenbou, Kindaichi, Shibata, Yamada, & Kindaichi). A subset of 78,801 words from this list was used in Edwards and Beckman (2008). Also, for English, which has more than five vowels, I collapsed together vowels that have similar coarticulatory effects. Specifically, I included both lax and tense vowels in each vowel category where the tense/lax contrast is relevant (for example, both /i/ and /I/ were included in the /i/ category) and I included all three low back vowels /ɑ, ʌ, ɔ/ in the /a/ category. The Mandarin sibilant frequencies were calculated by the author using the online LDC consortium on Mandarin lexicon (Huang, Bian, Wu, and McLemore, 1996). For all three languages, following Edwards and Beckman (2008), each CV type were calculated by dividing the occurrence of this CV sequence in the word-initial position by the total number of words in the corpus, and then calculated the log of this ratio. All the frequency calculations are tabulated in table 2.2.

Table 2.2 shows both the overall phoneme frequency and the CV frequency for every voiceless sibilant in each of the three languages. For the overall phoneme frequency, it is clear the English /s/ is twice as frequent as that of /ʃ/, whereas in Japanese /s/ and /ɕ/ are comparable in frequency, with /s/ being slightly more frequent than /ɕ/. In Mandarin, /s/ is lower in frequency compared with the other two fricatives. When consonant frequency was examined in the context of the following

		English		Japanese		Mandarin		
		/s/	/ʃ/	/s/	/ɕ/	/s/	/ʂ/	/ɕ/
Phoneme frequency		-3.1	-5.0	-2.3	-2.5	-3.9	-2.7	-2.8
CV frequency	/-a/	-4.7	-6.8	-3.5	-5.3	-5.1	-4.3	-4.0
	/-i/	-4.5	-6.3	–	-3.1	-5.2 (/sɹ/)	-4.1	-4.9 (/ʂɻ/)
	/-o/	-6.1	-7.0	-4.0	-4.1	-6.2	-5.1	-5.6
	/-e/	-4.6	-6.6	-3.6	-8.3	–	-9.6	-4.7
	/-u/	-5.9	-7.7	-3.9	-4.4	-5.1	-4.3 (/ɕy/)	-4.7

Table 2.2: Log phoneme frequency and CV frequency for the sibilant fricatives in English, Japanese and Mandarin Chinese.

vowel, the pattern generally agreed with the overall phoneme frequency distribution. That is, in English, /s/ is always more frequent than /ʃ/ in all vowel contexts. In Japanese, the CV frequency distribution is consistent with the overall phoneme frequency distribution except for the sequence /si/ which is phonotactically illegal. Similarly, in Mandarin, this is the case for the sequence /se/, which rarely occurs in the language.

2.5 Summary

To summarize, in terms of articulatory gestures, English /s/ and /ʃ/ mainly contrast in regards to where the major lingual constriction is made in the oral cavity, or the place of articulation; the tongue position for /s/ is further front than that for /ʃ/. The Japanese /s-ɕ/ pair, however, not only differ from each other in where the major constriction is, but also in how the constriction is made by the tongue. Specifically, the production of /ɕ/ involves a palatalized tongue posture which creates a long palatal channel between the mid-sagittal surface of the tongue and the upper roof of the mouth, whereas that of /s/ does not. For Mandarin Chinese, the distinction between /s/ and /ʂ/ mainly lies in the constriction place, with the major constriction

for /s/ being made further front than that for /ʂ/. /s/ and /ʂ/ also contrast with /ɕ/ in tongue posture, since the tongue is palatalized in producing /ɕ/ in Mandarin.

In addition to articulatory aspects, sibilant fricatives in the three languages differ in terms of their contrastive status in different vowel contexts, as well as phoneme frequencies. In English, /s/ and /ʃ/ have a robust distribution in front of all vowels. In Japanese and Mandarin, the sibilant fricatives only contrast in some vowel contexts. English /s/ and /ʃ/ differ in phoneme frequencies with /s/ being more frequent than /ʃ/ in word-initial position, whereas Japanese /s/ and /ɕ/ share similar phoneme frequencies. In Mandarin, /s/ has lower frequency counts than the other two sibilants. The effect of phonological distribution and phoneme frequencies are discussed later when necessary.

CHAPTER 3

ACOUSTIC PATTERNS OF ADULT VOICELESS SIBILANT FRICATIVE
PRODUCTIONS

3.1   Introduction

This chapter describes acoustic patterns in adult productions for each of the three
languages. Right after infants are born, they are immediately immersed in the ambi-
ent language environment. The way that adults speak will affect both how children
perceive the speech categories and how they produce them. Although the data in-
cluded here are not child-directed speech, and the link between adults' productions
and children's productions may not be so straightforward, children still will pay atten-
tion to both their primary care-takers' speech and to how people speak around them.
And when compared across languages, such influence is more apparent. Therefore, it
is important to describe the adult production patterns in different languages in order
to know the norms that children eventually target.

Because the voiceless sibilants contrast with each other in different articulatory
aspects in the three languages as shown in Chapter 2, it is hypothesized that the
acoustic realizations of the contrast in each language will be thus different. The goal of
this chapter, therefore, is to establish relevant acoustic parameters that can effectively
differentiate the sibilant contrasts in adult productions in each of the languages. More
importantly, these acoustic parameters will be used to describe the adult patterns and

compare them across languages. The adults' production patterns will then later serve as the baseline to be compared with children's productions in order to find out when children start to show language-specific tendencies in their productions.

3.2   Acoustics of voiceless sibilant fricatives

As shown in Chapter 2, although the fricatives in these three languages can be roughly equated across languages so that all three languages can be described as having /s/ and at least one postalveolar fricative, they differ from each other in subtle articulatory gestures. An important question is how to capture these articulatory contrasts using acoustic parameters. Previous research on languages such as English that have a place-of-articulation contrast for sibilant fricatives has focused on differentiating the two fricatives based on the spectral properties of the frication noise (Behrens and Blumstein, 1988; Hughes and Halle, 1956). The fricative /ʃ/ in English has a longer front cavity than /s/ both because of its more posterior place of articulation and also because of its characteristic lip rounding when preceding unrounded vowels. This difference in front cavity length results in more low-frequency energy for the /ʃ/ spectrum and more high-frequency energy for /s/ (Hughes and Halle, 1956; Stevens, 1998).

A commonly used method for examining the spectral properties of fricative noise is the spectral moments analysis, in which the power spectrum is treated as a probability distribution so that the statistical moments can be calculated (Forrest, Weismer, Milenkovic, and Dougall, 1988; Shadle and Mair, 1996; Jongman, Wayland, and Wong, 2000). The first spectral moment, M1, (the mean or "centroid" frequency) describes the weighted mean frequency in the fricative spectrum, and it works well to distinguish between /s/ and /ʃ/ in English, as shown in many previous studies (Forrest et al., 1988; Shadle and Mair, 1996; Jongman et al., 2000; Nittrouer et al., 1989). In

a spectrum with only one prominent mode, the frequency of the first moment should be negatively correlated with the length of the front resonating cavity, and thus can roughly describe where the constriction is made relative to the front end of the oral cavity.

The second spectral moment (standard deviation), or M2, describes how the fricative spectrum deviates from the mean frequency. It does not seem to be useful in distinguishing between the two sibilant fricatives of English since it is mainly used to differentiate between a flat diffuse spectral shape and a peaky, compact distribution. Instead, it has been found to be able to distinguish sibilants from nonsibilants (Jongman et al., 2000; Nissen and Fox, 2005). In addition, Stoel-Gammon, Williams, and Buder (1994) found that the second spectral moment is effective in differentiating Swedish /t/, which is lamino-dental, from the American-English /t/, which is apical-alveolar, with the former having a more diffuse spectral shape than the latter. This parameter may thus help to distinguish Japanese /s/ and /ɕ/, since according to Akamatsu (1997), Japanese /s/ is a lamino-alveolar/dental fricative and is less sibilant than American English /s/, and therefore is likely to have a more diffuse spectral shape.

There are two other moments that people have used before. One is the third spectral moment (skewness), which calculates how skewed the spectral shape is by subtracting the frequency range of the spectrum below the centroid from that above the centroid. In previous studies, M3 has also shown to be useful in distinguishing between /s/ and /ʃ/ in English, as it is correlated with place-of-articulation distinction. In general, /ʃ/ should have a positive value, indicating an energy concentration in the frequencies below the mean value, while /s/ should have a negative value, indicating a concentration of energy in the frequencies above the mean value. The other commonly used spectral moment is the fourth spectral moment (kurtosis) which

describes how much the spectral shape is different from a normal distribution. It has been found to be useful in differentiating sibilants from nonsibilants (Jongman et al., 2000), as these differences result in changes in peakiness of the spectral shape.

Spectral moments analysis has also been extended to classify English-speaking children's fricative productions. For example, Nittrouer (1995) used it to compare the productions of /s/ and /ʃ/ in children aged 3 to 7 and adults, and found that the difference in M1 between /s/ and /ʃ/ is larger for adults than for children. Similarly, Nissen and Fox (2005) used spectral moments to classify the fricative productions in English-speaking children of age 3 to 6. Their results show that adults and the 5 year-old group display significant differences between the two fricative categories in M1, whereas in younger age groups such differences can not be found.

It has to be noted that moments analysis is not the only type of analysis that has been used in the past to describe fricative spectra. Shadle and colleagues (Shadle and Scully, 1995; Shadle and Mair, 1996; Jesus and Shadle, 2002) have developed other acoustic parameters, such as frequency of the highest amplitude peak, as alternatives to the first moment for place contrast. However, since moments analysis has been used more extensively, especially in describing children's productions, and has been shown to work well, I will use it for analyzing fricative spectra in this dissertation. (A more systematic comparison between moments analysis and other types of analysis in different scales is planned for future research.)

In addition to fricative-internal spectral parameters calculated over the fricative duration, Stevens et al. (2004) proposed a transitional parameter, onset F2 frequency, to capture the tongue posture difference between /ɕ/, /s/ and /ʂ/ in Mandarin Chinese. More specifically, onset F2 frequency refers to the second formant frequency taken at the onset of the following vowel, and is considered to correlate negatively

with the length of the back cavity immediately after the release of a coronal constriction (Nittrouer et al., 1989; Funatsu, 1995; Tsurutani, 2004). This is because the narrowest constrictions starting at the alveolopalatal position are anterior to where the lowest front cavity resonance crosses the lowest back cavity resonance. Since the alveolopalatal /ɕ/ sound has the longest channel, and thus shortest back cavity, the onset F2 frequency of /ɕ/ should be higher than that for either /s/ or /ʂ/. Therefore, the acoustics of the palatalization contrast is realized by both fricative-internal and fricative-vowel transitional characteristics in the speech signal. Similarly, Funatsu (1995) examined the acoustics of the Japanese contrast between /s/ and /ɕ/, and the contrast in Russian among /s/, /sj/, and /ʃ/, and found that the onset F2 frequency together with the main peak frequency in the fricative noise are sufficient to describe the fricative contrasts in both languages. It is clear from Funatsu (1995) that the palatalization contrast in Japanese can be distinguished jointly by both parameters, with /s/ having lower onset F2 frequency and higher spectral peak frequency than /ɕ/. The Russian fricative contrast, however, exhibits a different pattern, in that the onset F2 frequency alone can distinguish /s, ʃ/ from /sj/ (the palatalization contrast), and the spectral peak frequency can distinguish /s, sj/ from /ʃ/ (the place distinction).

Table 3.1 lists the five acoustic parameters that are used in the acoustic analysis of fricative productions along with their acoustic definitions, possible articulatory interpretations, and a record of the previous studies that have used them in describing either adults' or children's productions. It should be noted that this dissertation does not include a component on the articulation of fricatives, and therefore all the articulatory interpretations of the acoustic parameters were based on previous studies.

	Acoustic parameter	Definition	Tentative articulatory interpretation	Used in previous studies
Fricative spectrum moments	M1 (Centroid)	The weighted mean frequency	Negatively correlates with the length of the front resonating cavity.	Nittrouer et al. (1989); Nittrouer (1995); Tsurutani (2004); Nissen and Fox (2005); Forrest et al. (1988); Shadle and Mair (1996); Jongman et al. (2000)
	M2 (Standard Deviation)	Average squared distance from the centroid	Differentiates tongue posture between apical/alveolar and laminal/dental.	Nittrouer (1995); Nissen and Fox (2005); Forrest et al. (1988); Jongman et al. (2000); Stoel-Gammon et al. (1994)
	M3 (Skewness)	How skewed the spectral shape is (calculated by subtracting the spectrum energy below the centroid from that above the centroid)	Negatively correlates with the length of the front resonating cavity.	Nittrouer (1995); Nissen and Fox (2005); Forrest et al. (1988); Jongman et al. (2000)
	M4 (Kurtosis)	How much the shape of the spectrum around the center of gravity is different from a Gaussian shape	Differentiates tongue posture between apical and laminal.	Nittrouer (1995); Nissen and Fox (2005); Forrest et al. (1988); Jongman et al. (2000); Stoel-Gammon et al. (1994)
CV transitions	Onset F2 frequency	F2 frequency at the onset of the following vowel	Negatively correlates with the length of the back resonating cavity.	Nittrouer (1995); Funatsu (1995); Stevens et al. (2004); Tsurutani (2004)

Table 3.1: Summary of parameters for acoustic analyses.

3.3 Methods

3.3.1 Participants

In order to describe adult production norms first, a series of production experiments were conducted. More specifically, for each language, the participants include 10 adults aged from 18 to 30 years. Gender was balanced for speakers of each language. All the participants were recorded in their home countries. That is, all English subjects were recruited in Columbus, Ohio, US; all Japanese subjects were recruited in Tokyo, Japan; all Mandarin subjects were recruited in Songyuan, China, to ensure dialect homogeneity. All speakers tested have normal hearing and had passed a hearing screening using otoacoustic emissions at 2000, 3000, 4000, and 5000 Hz. No adults tested have reported histories of speech, language, or hearing problems.

3.3.2 Materials

All participants were recorded while they were engaged in a word-repetition task. The materials were word-initial voiceless fricatives preceding vowels in words that are familiar to children, since the procedure for testing children was exactly the same, as shown in Chapter 6. The vowels are roughly grouped into a set of five categories /i e a o u/, as Japanese has only these five vowels. For English, vowels that have similar coarticulatory effects were collapsed together. Specifically, both lax and tense vowels were included in each vowel category where the tense/lax contrast is relevant (for example, both /i/ and /ɪ/ were included in the /i/ category) and all three low back vowels /a/,/ɔ/,/ʌ/ were included in the /a/ category, and the vowels /ɛ/ and /e/ were collapsed into the /e/ category. Similarly, for Mandarin, the vowels [i], [ɻ̩] and [ɹ̩] were counted into the /i/ category, and /u/ after /s/ or /ʂ/ and /y/ after /ɕ/

were classified into the /u/ category. All these word-initial CV sequences were elicited within familiar words in all three languages, accompanied by pictures. There were approximately three target words for each CV sequence, though not all of the CV sequences could be elicited because of phonotactic constraints. For example, */si/ is unattested in Japanese. Also /ɕe/ is attested only marginally in Japanese, primarily in recent loan words from languages such as English. In English, only two words containing /ʃu/ were elicited because there are few words containing this sequence that are familiar to young children. A complete list of words for each language is listed in Table 3.2, 3.3 and 3.5.

For all three languages, the stimulus items for the word-repetition task were spoken by an adult female native speaker in a child-directed speech register. The speaker was familiar with the purpose of the task. The fricative-initial words were recorded in a randomized list along with other words that began with other lingual obstruents (stops and affricates). For each word type, three tokens were presented to adults. Two tokens that were perceived with at least 80 percent accuracy by the five adult native speakers were selected for use with the actual testing on both adults and the children.

3.3.3   Procedure

Participants were tested in a quiet room. Each stimulus item was played out over speakers connected to a computer sound card and the adult participants were asked to repeat each item as they heard it. Each trial item consisted of a picture and the associated sound file, which were presented simultaneously to the participant over a laptop with a 14-inch screen using a program written specifically for this purpose. The computer program included an on-screen VU meter to help the participants monitor their volume and a picture of a duck walking up a ladder on the left side of the screen

Target fricatives	Target vowels	Words	WorldBet transcriptions
s (/s/)	A	sun	sˆn
		sauce	sas
		soccer	sa.k3
	E	same	sem
		safe	sef
		seven	sE.vIn
	I	seashore	si.Sor
		seal	sil
		sister	sIs.t3
	O	soldier	sol.dz3
		soak	sok
		sodas	so.d&z
	U	super	su.p3
		suitcase	sut.kes
		soup	sup
S (/ʃ/)	A	shovel	Sˆ.v&l
		shark	Sark
		shop	Sap
	E	shell	Sel
		shepherd	SE.p3d
		shape	Sep
	I	sheep	Sip
		ship	SIp
		shield	Si.=ld
	O	shore	Sor
		show	So
		shoulder	Sol.d3
	U	chute	Sut
		sugar	SU.g3
		shoe	Su

Table 3.2: Word list of the word repetition task for English-speaking children and adults.

Target fricatives	Target vowels	WorldBet transcriptions	Gloss
s (/s/)	A	sakana	fish
		sakura	cherry blossom
		saru	monkey
	E	semi	cicada
		senaka	back
		seNqsee	teacher
	O	sokkusu	socks
		sooseedZi	sausage
		sora	sky
	U	sudzume	sparrow
		suika	watermelon
		suna	sand
S (/ɕ/)	A	Sawaa	shower
		SamodZi	rice paddle
		Sampuu	shampoo
	I	SiNqkaNqseNq	bullet train
		Siisoo	seesaw
		Simauma	zebra
	O	SoobooSa	fire engine
		SokupaNq	bread
		Sooju	soy sauce
	U	Suukuriimu	creme puff
		Suumai	Chinese dumpling
		Suudzu	shoes

Table 3.3: Word list of the word repetition task for Japanese-speaking children and adults.

Target fricatives	Target vowels	WorldBet transcriptions	Gloss
s (/s/)	A	san1.jiao3 sa1.niao4 san3 sa3.le	triangle to pee umbrella spill over
	E	sei1.ya2	(food) stuck between the teeth
	I	sI1.ji1 sI4.miao4 sI4.ge sI1	driver temple four silk
	O	song1 song1.jin3.dai4 song1.le song1.shu3	pine tree elastic band loose squirrel
	U	sun1.wu4.kong1 su4.liao4 sun1.zi sun3	Monkey King plastics grandson bamboo shoots
S (/ʂ/)	A	shan1 shan1.yang2 sha1.fa1 shan4.zi	mountain goat sofa fan
	I	shI4.zi shI1.zi shI2.zir3 shI2	lion tomato pebble ten
	O	shou3 shou3.juan4 shou4.zi shou3.tao4	hand handkerchief a thin person mittens
	U	shu4 shu1.bao1 shu1.cai4 shu1.shu	tree bookbag vegetables uncle

Table 3.4: Word list of the word repetition task for Songyuan-speaking children and adults.

Target fricatives	Target vowels	WorldBet transcriptions	Gloss
c} (/ɕ/)	A	xia4.yu3 xiang1.jiao1 xia1 xiang1.zi	rain banana shrimp trunk
	E	xian4	elephant
	I	xi4 xi1.gua1 xing1.xing xi3.shou3	thin watermelon stars hand washing
	O	xiong2 xiu1.xi3 xiong2.mao1 xiu4.zi	bear rest panda sleeve
	U	xv3.duo1	many

Table 3.5: Word list of the word repetition task for Songyuan-speaking children and adults. (Continued)

to provide visual feedback to the participants about their progress in completing the task. There was a practice session prior to administration of the experimental task.

All participants' responses were digitally recorded onto a Marantz PMD660 flash card recorder. All the recordings were done using AKG C5900M condenser vocal microphone, positioned using a metal mic stand on a desk surface in front of them, and the average distance between the speaker's mouth and the microphone was about 30 cm. The angle between the mouth of the participants and the microphone was about 45 degrees. All the data were recorded using a sampling rate of 44,100 Hz at 16 bits.

3.3.4 Segmentation

Praat (Boersma and Weenink, 2005) was used to segment words from larger sound files, and to identify fricative boundaries. The beginning of the fricative is defined as the beginning of the frication, identified both by a clear increase in the frication noise amplitude in the waveform and by the presence/occurrence of white noise in a frequency band above 1000 Hz in the spectrogram. The end of the frication is defined as the beginning of the following vowel, and in particular, the first zero crossing of an upswing pitch cycle of the first periodic glottal pulse of the vowel. To ensure maximal consistency across tokens, the dynamic range in spectrogram view in Praat was set consistently to be 40dB, and the view range is from 0 Hz to 8000 Hz for adults, and from 0 Hz to 11000 Hz for children. An example is shown in Figure 3.2.
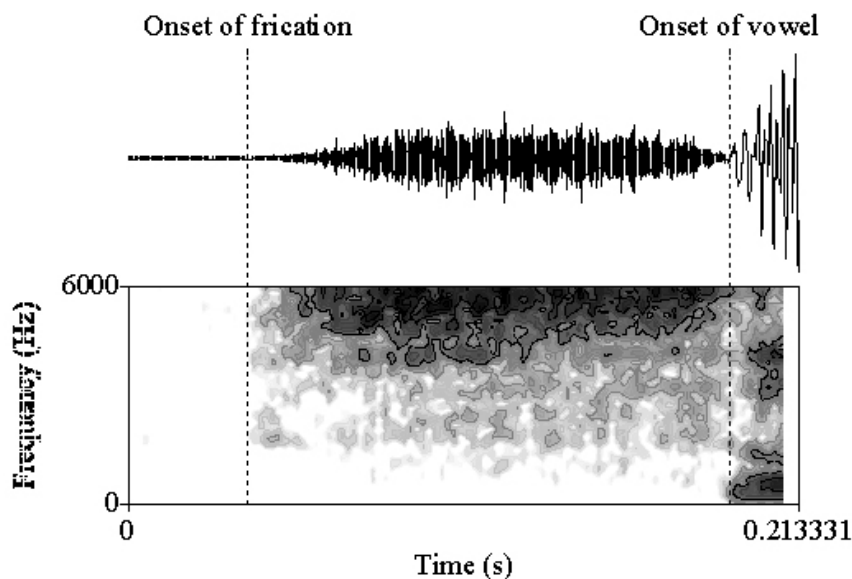
Figure 3.1: Segmentation of a /ɕa/ sequence at the beginning of the word /sõ/ "pine" produced by a Mandarin female speaker.

For the spectral moments analysis, an FFT spectrum was extracted over a 40 ms Hamming window centered around the midpoint of the fricative noise. The middle 40 ms window was chosen because it is the most steady portion of the fricative noise and is least likely to be influenced by amplitude buildup at the beginning of the fricative or the transitional change into the following vowel. Onset F2 frequency was measured at the end of the fricative noise. The setting in Praat that we used to estimate the onset F2 was an LPC analysis specified for 5 formants (10 coefficients) calculated over a range from 0 to 5500 Hz for adults and from 0 to 7000 Hz for children. The window length was 0.025 ms. All calculations were made without pre-emphasis.
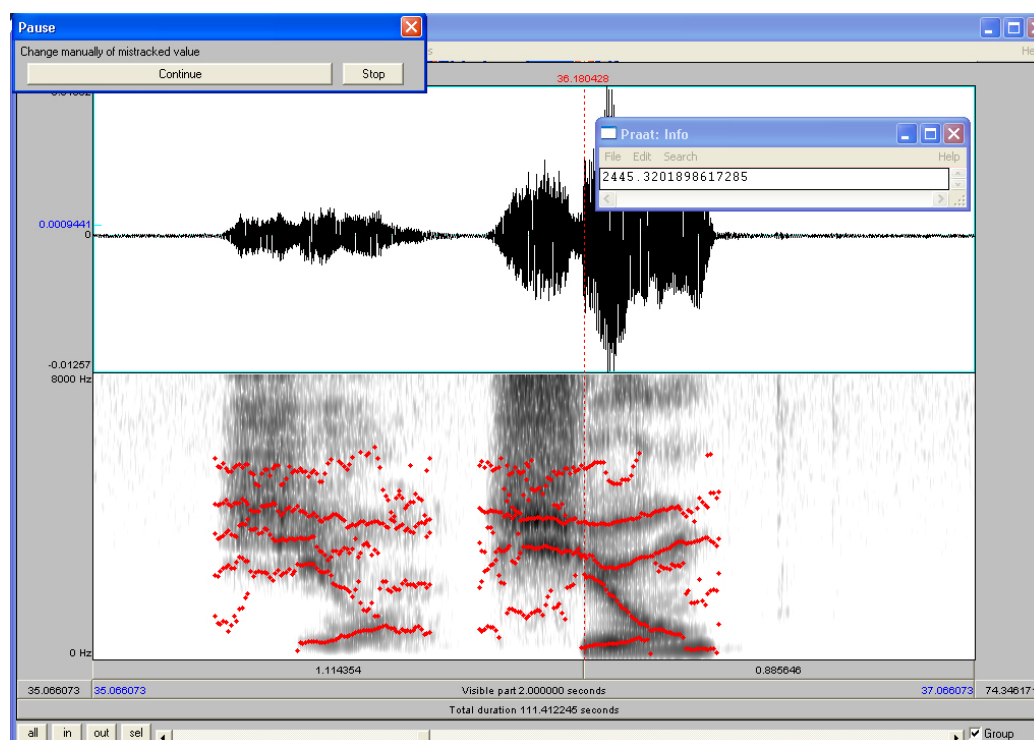
Figure 3.2: Snapshot of semi-automatic extraction of onset F2 frequency using Praat.

After the fricative edges were marked, a Praat script was used to automatically extract the four moment values within all aligned tokens over a 40 ms window centered

34

in the middle of the fricative noise segments. The four moments were calculated over the frequency range between 1000 Hz and 22500 Hz to eliminate the effect of low frequency noise. The extraction of the transitional parameter, the onset F2 frequency, however, is semi-automatic, in that the script paused at every token to be extracted with the automatic value shown in the screen along with the spectrogram to enable hand corrections of mistracked tokens. A snapshot of the script execution when extracting onset F2 frequency values is shown in Figure 3.2.

3.4   Results of acoustic analysis

3.4.1   Averaged spectra

Since four out of the five acoustic parameters used were calculated from the fricative spectrum, it will help to examine the overall spectral shapes for different fricative categories and for different languages in order to get an idea of which acoustic parameters might be useful for different languages.

The procedure for averaging spectra is as follows: As mentioned earlier, for each fricative token (i.e. individual word produced by each speaker), an FFT spectrum was calculated over the middle 40 ms window of the frication noise. The resulting spectrum was converted to a Praat "long term average spectrum" (Ltas) object with a 10.77 Hz bin size. These Ltas objects were then averaged over all the tokens produced by five female speakers and all those by the five male speakers of each language to produce two averaged spectra per language. Figure 3.3 shows these grand averaged spectra for the adult speakers of the two genders in the three languages. Because of the difficulty of inputing phonetic symbols in the graphical display, all the figures in this dissertation will adopt the worldBet symbol of '/S/' for the sound /ʃ/ in English,

as well as /ʂ/ in Mandarin, and the symbol of '/c}/' to represent the alveolopalatal /ɕ/ in both Mandarin and Japanese.

The reason to separate out speakers of the two genders is that their fricative spectra may differ because females have shorter vocal tracts, similar to the effect of higher formants that has been shown in vowel productions. For sibilants, where the resonating cavity is the tube in front of the major lingual constriction instead of the whole vocal tract, this effect of biological difference is not as straightforward as in the case of vowel production. However, the front resonating cavity of female speakers is still expected to be shorter in proportion to the length of the whole vocal tract, yielding a higher spectral peak in the spectrum, and this actually have been shown to be true in several studies (Jongman et al., 2000; Jesus and Shadle, 2002).

In English, the averaged spectrum of /s/ for female speakers (upper left panel of Figure 3.3) has a well-defined peak around 10,000 Hz. By contrast, the averaged spectrum of target /ʃ/ seems to have two peaks, with the major peak occurring around 3000 Hz or 4000 Hz. This major difference in the dominant frequency range should reflect in M1, which calculates the centroid frequency in the fricative spectrum. Based on the averaged spectra, it is expected that English /s/ tokens should have higher M1 values than /ʃ/ tokens. Another thing to note is that the spectral peaks of both target fricatives generally shift downward for males as compared with females. Also the two fricative spectra are not as far apart in males' productions. This is in accordance with the prediction that men and women would realize the fricatives differently in acoustics because of different length of the front resonating cavity.

The distributions of both target fricatives in Japanese (the middle panel in Figure 3.3)) are flatter, especially for /s/, as compared with the two English fricatives. M2, which describes the variance of the spectrum, should be able to capture the more diffused spectral distribution of /s/ in Japanese. Meanwhile, the two Japanese
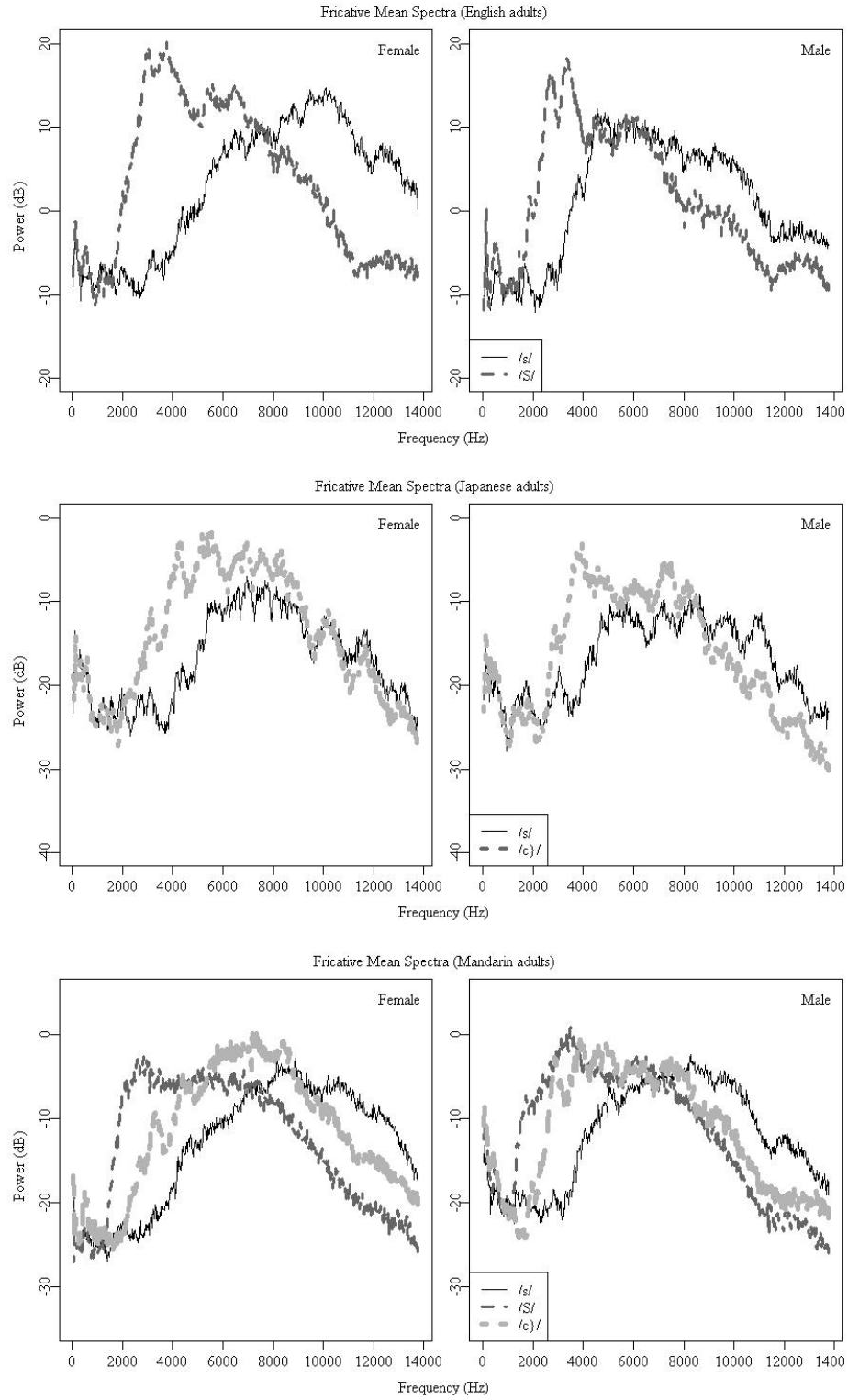
Figure 3.3: Averaged spectra of contrastive fricatives in adults' productions for the three languages.

fricatives are not as far apart as the English pair, as shown by the great overlap in M1. This leads to the prediction that the M1 difference between the two targets in Japanese may not be as big as that between the two English fricatives. The two genders of Japanese speakers do not seem to differ as much as the English speakers as the averaged spectra pattern similarly for the two genders, although the separation between the two target fricatives is still smaller for male speakers than for female speakers owing to biological differences between the two sexes.

In the case of Mandarin (the bottom panel), the major peaks of /s/ and /ʂ/ are separated from each other in a way that is similar to the English pattern, while /ɕ/ fits in the middle frequency range between /s/ and /ʂ/. These differences in the peak frequency of the distribution can also be described using M1, with the predicted M1 value being the highest for /s/, lowest for /ʂ/, and intermediate for /ɕ/. The spectra of /ʂ/ and /ɕ/ are less separated from each other in the male speakers' productions. This reflects a phenomenon called 'feminine accent' in Chinese sociolinguistic literature, which describes the fronting of alveolopalatals in females' speech (Cao, 1986; Hu, 1991). The fronting is interpreted as social modification on the females' part in addition to the natural 'fronting' caused by genetic differences between the two sexes.

3.4.2 Distributions of the three acoustic parameters

The averaged spectra depicted in 3.3 describe the distributions of the raw fricative spectrum. This section further explores the distributions of the five acoustic parameters for each of the target fricatives in the three languages, separated by gender. Among the five acoustic parameters, M1 is correlated with the frequency of the major spectral energy concentration. M2, M3 and M4 have to do with the shapes of the spectra, and the onset F2 describes the fricative-vowel transition in F2 that is correlated with the length of the back cavity.

Figure 3.4 plots the mean values for the five acoustic parameters for the 10 adult speakers of English. The mean values were calculated by gender and by vowel context. In the figure, the first row plots the mean M1 frequencies for the two genders in each vowel context, with the error bar indicating 1 standard deviation above and below the mean. It shows that both males and females have very distinct M1 values for /s/ and /ʃ/. The rounding of the vowel /u/ and /o/ lowers the centroid frequencies for both genders, which is expected given that the protrusion of the lips in producing rounded vowels lengthens the front resonating cavity. Another thing to notice is the different pattern the two genders exhibit: both males and females have similar centroid frequencies for the /ʃ/ sound, with females' centroid values being slightly higher than those of males. Strikingly, the centroid values of the /s/ sound produced by female speakers is much higher than those produced by male speakers. Moreover, the variability of /s/ for female speakers is bigger than that for male speakers, and the amount of variability is similar across the vowel contexts.

The second row of Figure 3.4 shows the mean M2 frequencies for the two genders in different vocalic contexts. For both genders, the distribution of the M2 does not differ much in different vowel contexts. Moreover, unlike the gender-differentiating pattern in M1 dimension, males tend to show bigger differences in M2 for the /s-ʃ/ contrast than females. Relating these high M2 values in the male speakers' productions with the flat spectral shape in male speakers in the top right panel of Figure 3.3 suggests a more laminodental quality in men's production of /s/.

The third row in Figure 3.4 shows the mean values of M3 for the two targets. Notably, the separation between /s/ and /ʃ/ is less clear in this dimension. Even if the mean M3 values for /ʃ/ are generally higher than those for /s/, the variability is too big to make the distinction robust. The pattern of M4 in the fourth row of the Figure is even less clear because of the huge variability.

The fifth row in the graph is about the differentiation between the two fricatives in the dimension of the onset F2. Different from the above two spectral measures, onset F2 frequencies vary more according to the vowel context. That is front vowels /i/ and /e/ have higher onset F2 frequency values than back vowels /a/ and /o/, which is expected since onset F2 is a measure that indexes the length of the back cavity. And because it is taken at the end of the frication and the onset of the following vowel, it is affected greatly by the tongue advancement of the vowel as well. The value of onset F2 for the vowel /u/, however, is higher than expected, which suggests that the American English /u/ is slightly fronted.

A similar plot of the Japanese adults' /s-ɕ/ productions is presented in Figure 3.5. One thing that is immediately noticeable is that the separation of the two fricatives is much less for both genders in the M1 dimension compared with that of English speakers. Further, the mean M1 values of /ɕ/ for females are higher than those for males, which is predictable given the biological makeup of the the genders, and which is also similar to the /ʃ/ pattern for English adult speakers between the two genders. For both genders, the M1 values for /s/ are higher than those for /ɕ/, suggesting that /s/ is produced with a fronter constriction in the oral cavity than /ɕ/. In the M2 dimension, which is the 2nd row in Figure 3.5, female speakers do not show systematic variation for the /s-ɕ/ distinction across vowel contexts. Male speakers, however, show consistently higher values for /s/ than for /ɕ/, and the separation between the two target fricatives is bigger than that of female speakers. This suggests that Japanese male speakers may produce more laminal /s/s than females. In the M3 and M4 dimensions, the variabilities are again very great, and the separation between the two targets is not very clear. Further, in the M4 dimension, female speakers show slightly higher M4 in target /s/, whereas male speakers have the opposite pattern – the M4 values for /ɕ/ are higher. In the fifth row of the figure, similar to the English

40
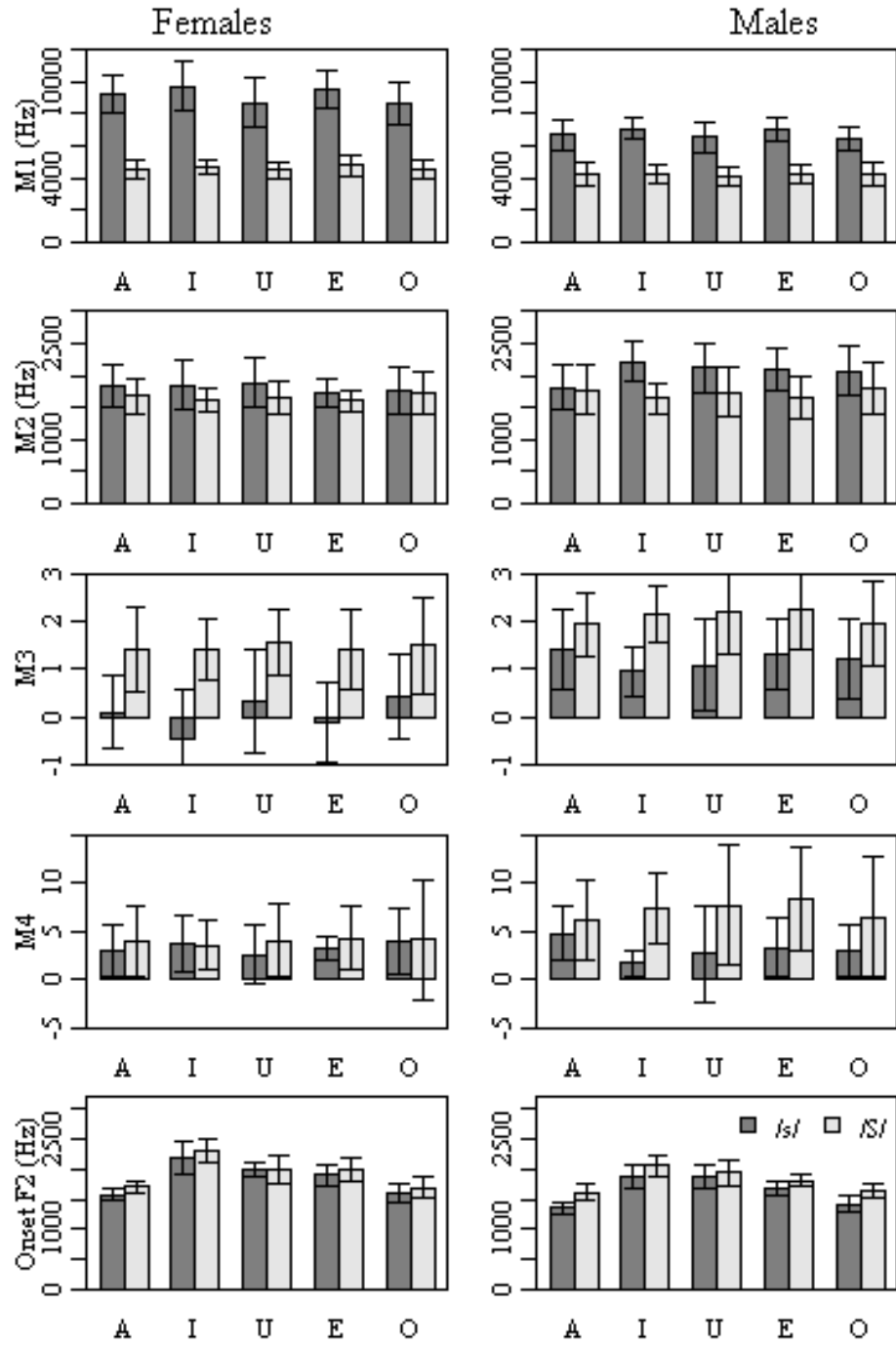
Figure 3.4: Acoustic descriptions of English adults' /s-ʃ/ contrast using three acoustic parameters: the first and the second moment in moments analysis (centroid frequency and the standard deviation respectively) as well as the onset F2 frequency.

pattern, onset F2 frequency varies both in terms of fricative category and in different vocalic context. Female speakers show better distinction in /s-ɕ/ contrast in the onset F2 dimension than for male speakers.

For Mandarin, the directions of the two sets of contrasts being made and the interactions of acoustics parameters with gender can be viewed clearly in Figure 3.6. For example, in the M1 dimension, all three categories are well separated for both genders, with /s/ having the highest M1, and /ʂ/ the lowest, and /ɕ/ in the middle. This is exactly in accordance with the prediction, since /s/ is produced farthest front in the oral cavity and /ʂ/ is the farthest back. Also there is not much difference across vowel contexts except in females' productions before vowel /u/, which is likely to be caused by anticipatory rounding effect. One thing to note is that the M1 values for the /s/ and /ʂ/ of females are slightly higher than those for male speakers, which can be attributed to the different vocal tract lengths. The much higher M1 values of /ɕ/ for females, however, cannot be explained by biological differences alone. The higher M1 values here in females' /ɕ/ productions suggest a fronter constriction than that of males, and are reminiscent of the impressionistic descriptions in the Chinese sociolinguistics literature which describes that females front or dentalize their alveolopalatals to show their female identity. Another thing to note in this figure is that for the onset F2 dimension, /ɕ/ distinguishes itself from the other two categories with higher onset F2 frequency values for both genders. This is consistent with articulatory interpretations since the long palatal channel formed by the tongue blade and the upper roof of the oral cavity in producing the /ɕ/ sound shortens the length of the back cavity, giving rise to the high onset F2 values. There is no clear pattern that can be found in the M2, M3 and M4 dimensions for Mandarin speakers.
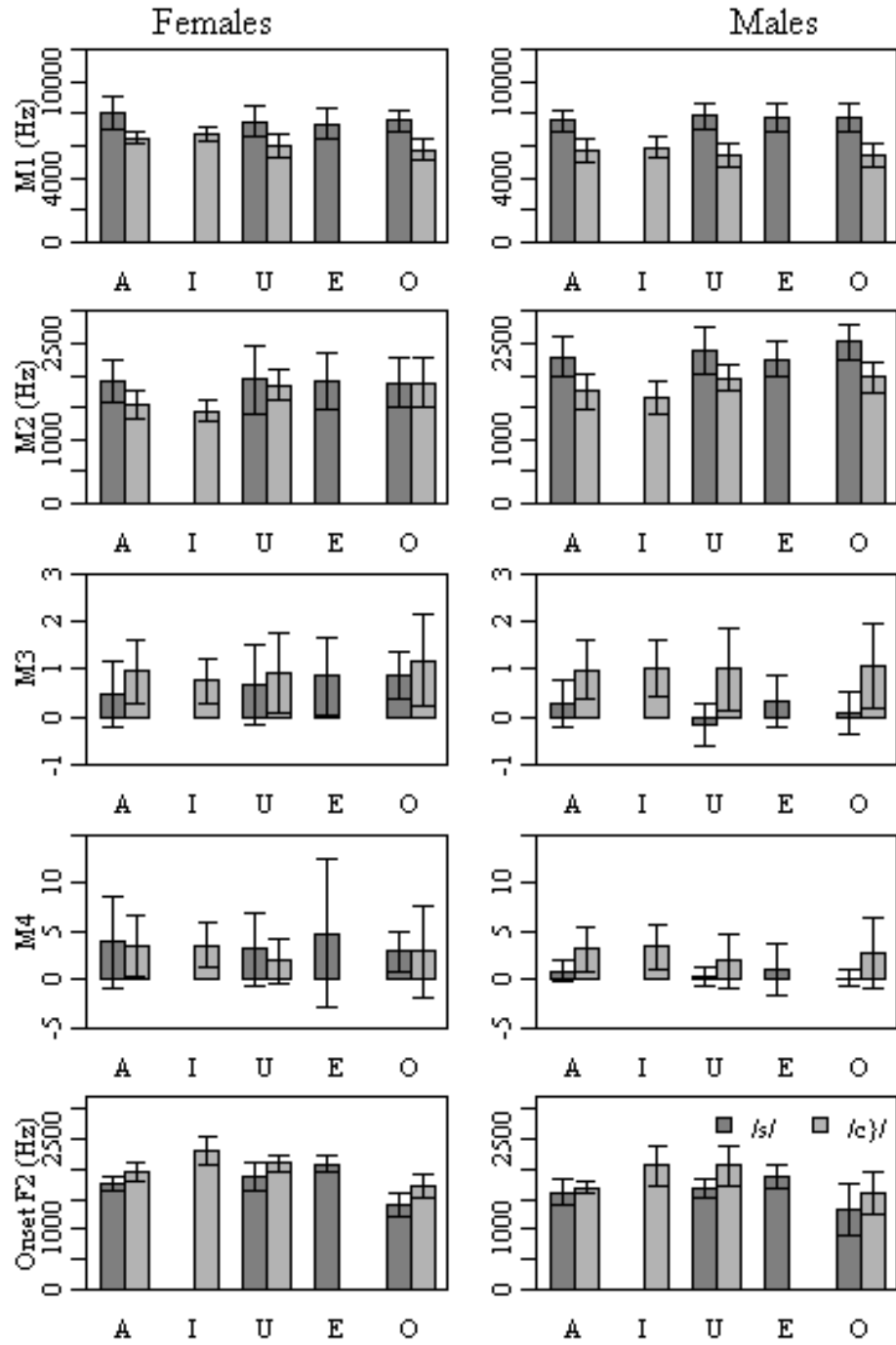
Figure 3.5: Acoustic descriptions of Japanese adults' /s-ɕ/ contrast using three acoustic parameters: the first and the second moment in moments analysis (centroid frequency and the standard deviation respectively) as well as the onset F2 frequency.
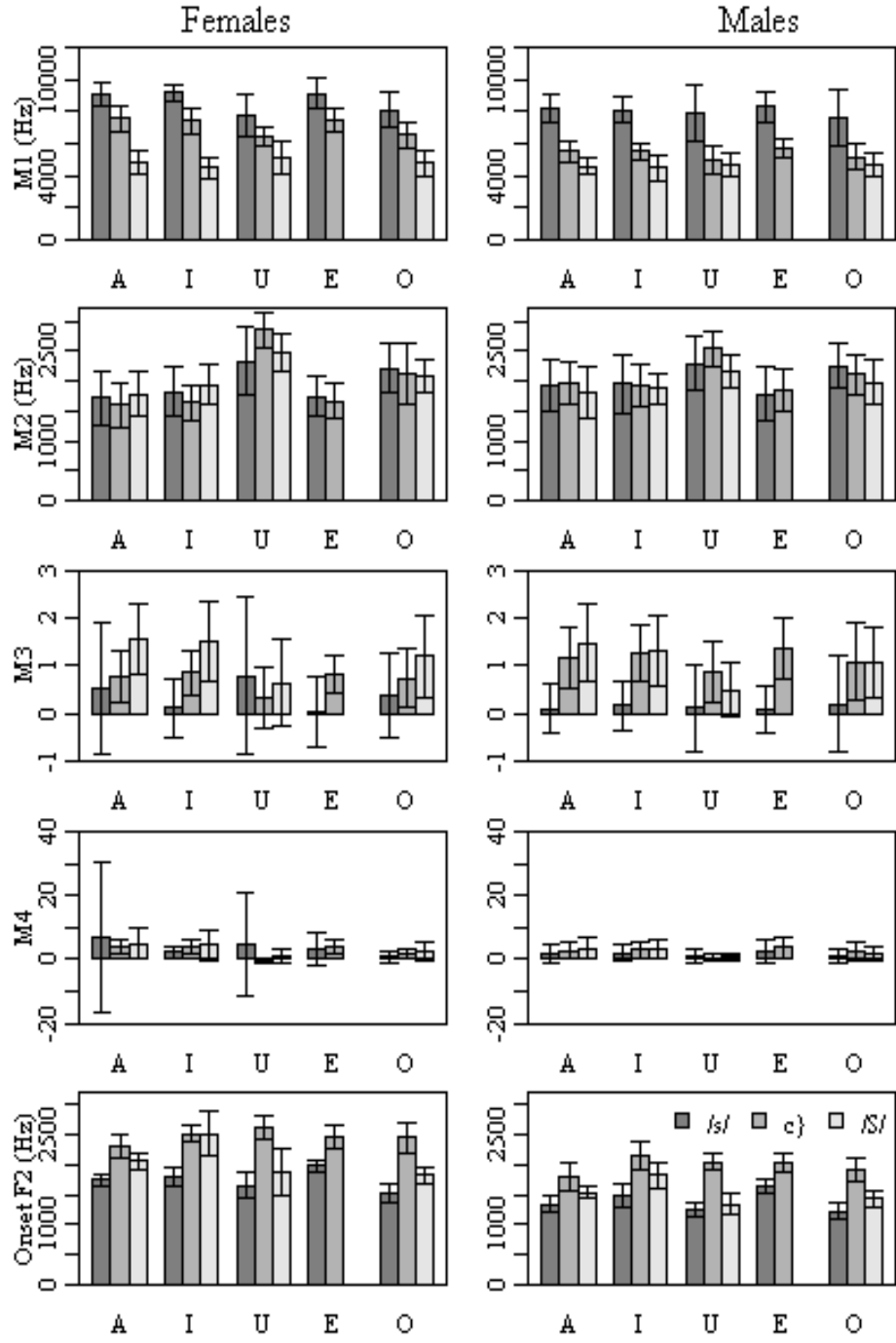
Figure 3.6: Acoustic descriptions of Mandarin adults' /s-ɕ-ʂ/ contrast using three acoustic parameters: the first and the second moment in moments analysis (centroid frequency and the standard deviation respectively) as well as the onset F2 frequency.

3.5 Statistical results

In order to quantify the observed patterns in Figure 3.3 - Figure 3.6, for each of the three languages a mixed effects logistic regression model (Jaeger, 2008; Baayen, Davidson, and Bates, 2008) was applied to determine which acoustic parameters are effective in predicting the fricative categories by specifying individual speakers as a random variable nested below each of the acoustic parameters. A mixed logit model is a type of Generalized Linear Mixed Effects model that has a linkage of logit function (Jaeger, 2008), and the logit function is ideal for binomial outcomes, such as in the case of English and Japanese which both have two contrastive fricatives. The Mixed effects model allows evaluations of the effects of interest (fixed effects) and the effects of manipulation such as subjects and items (random effects). Unlike the regular logistic regression, it has the advantage of dealing with variability that is not from a homogeneous source, but rather clustered variance coming from individual speakers or items that randomly fluctuate. The mixed effects model is also called the Hierarchical Linear Regression model since it admits covariance nested inside subjects or other grouping factors and are organized hierarchically.

For languages that have a two-way contrast in fricatives, such as English and Japanese, the dependent variables were the two target fricatives, and the independent fixed effects variables were M1, M2, onset F2, gender, and the interactions between gender and each of the three acoustic parameters. The reason to exclude M3 and M4 in the modeling is that these two parameters introduced collinearity into the overall model, since M3 correlates with M1, and M2 correlates with M4. Based on the patterns in Figure 3.4 through Figure 3.6, M1 and M2 are better acoustic correlates of the fricative contrast in the three languages, and have clearer and better defined

articulatory interpretations. Therefore, they are used together with onset F2 in the statistical analysis in this chapter.

The independent random effects variables are the individual subjects nested under each of the explanatory variables, that is, under each of the acoustic parameters as well as gender. This hierarchy of subjects nested within each of the explanatory variables allowed me to evaluate the effects of different acoustic parameters simultaneously after controlling for the variability introduced by individual subjects. Having gender as one of the explanatory variables allows me to evaluate whether the genetic differences between the two sexes are significant enough to predict the two fricative categories. Including the interaction of gender and other acoustic parameters could evaluate whether the two genders utilize the same acoustic parameters differently or not in realizing the two fricative targets. All the statistical analysis in this dissertation is done using R (R Development Core Team, 2007), an open source software for statistical computing.

3.5.1   English

The statistical results of the mixed effects logistic regression for English show that M1 alone is able to explain 100% of the variability, which makes fitting the above mentioned model impossible. In order words, the difference in M1 between the two fricative target is so salient that it alone is sufficient to explain the contrast between /s/ and /ʃ/ in English, and no other acoustic parameters are needed.

3.5.2   Japanese

Table 3.6 describes the results for the logistic regression of Japanese productions, with the dependent variables being /s/, which is coded as 0, and /ɕ/, which is coded as 1.

46

Formula: targetC ~ (m1 + m2 + onsetF2) * gender + (m1 + m2 + onsetF2 + gender | subj)				
Predictor	Coefficient	SE	Wald Z	p
Intercept	12.66	3.98	3.18	**0.001**
M1 (Centroid)	-0.0027	0.0005	-4.790	**<0.001**
M2 (Standard deviation)	-0.0011	0.0008	-1.325	0.185
Onset F2	0.0042	0.0012	3.432	**<0.001**
Gender (Male)	37.7982	14.5568	2.597	**0.0094**
M1 & Gender (Male)	-0.0022	0.0015	-1.501	0.1332
M2 & Gender (Male)	-0.0074	0.0039	-1.912	0.0559
Onset F2 & Gender(Male)	-0.0044	0.0022	-1.979	**0.0478**

Table 3.6: Summary of the mixed effects logistic regression model for Japanese adult speakers' productions.

It is clear from the table that the three acoustic parameters, M1, M2 and Onset F2 are significant in predicting the fricative categories in Japanese. The negative coefficient in the table for M1 indicates a negative correlation with the fricative /ɕ/, which is expected since /ɕ/ should have lower M1 than /s/ with the major constriction of former being slightly further back than that of the latter. The positive coefficient of onset F2 suggests a positive relationship between onset F2 values and the category /ɕ/. This is in accordance with the prediction since onset F2 indexes the length of the back cavity, and /ɕ/ has much shorter back cavity compared with /s/, owing to its long palatal channel. Comparing these two parameters that were shown to be significantly correlated with the productions of the two fricatives, M1 has more explanatory power than onset F2 as the absolute value of the coefficient of M1 is higher. In addition, gender interacts with M2 and onset F2 marginally in predicting the fricative categories. More specifically, male speakers are associated with lower M2 values and lower onset F2 in the /ɕ/ category than female speakers, as evidenced by the negative coefficients in the regression output.

3.5.3 Mandarin

Mandarin Chinese has a three-way contrast among /s/, /ʂ/ and /ɕ/. These three
fricatives contrast with each other in two articulatory aspects; alveolar /s/ contrasts
with the other post-alveolar fricatives place of articulation, with the major constric-
tion for /s/ farther front than that of the other two, /ɕ/ contrasts with /s/ and /ʂ/
in the tongue posture, as the production of /ɕ/ involves a palatalized tongue shape
whereas the other two sounds do not.

Since the logistic regression model only works on a dependent variable that has
two outcomes, two mixed effects logit models were fit in the Mandarin adults' data,
to evaluate the two articulatory aspects separately. One model is to evaluate the
place contrast by having /s/ as one category (the anterior category), and /ʂ/ as the
other category (the posterior category). The second model is for the palatalization
contrast, where one category is /ɕ/ (the palatalized category, which is coded as 1)
and the other category is /s/ and /ʂ/ together (the nonpalatalized category, which is
coded as 0).

Formula: posture ~ (m1 + m2 + onsetF2) * gender + (m1 + m2 + onsetF2 + gender | subj)				
Predictor	Coefficient	SE	Wald Z	p
Intercept	-0.0204	2.403	-8.514	**p<0.001**
M1 (Centroid)	0.0007	0.0001	6.041	**p<0.001**
M2 (Standard deviation)	-0.0004	0.0003	-1.135	0.2564
Onset F2	0.0068	0.0007	9.952	**p<0.001**
Gender (Male)	1.747	3.711	0.471	0.6378
M1 & Gender (Male)	-0.0008	0.0002	-3.964	**p<0.001**
M2 & Gender (Male)	0.0020	0.0007	2.742	**0.006**
Onset F2 & Gender (Male)	0.0019	0.0131	1.508	0.131

Table 3.7: Summary of the mixed effects logistic regression model for the palataliza-
tion contrast (/ɕ/ vs. /s,ʂ/) of Mandarin adult speakers' productions.

The results for the first model are similar to the ones in English, where M1 by itself is sufficient to predict the anterior category from the posterior category, with no other acoustic parameters being necessary. Thus, no output was generated in R for this model. For the second model, the statistical results are listed in Table 3.7. It is clear from the table that, for the palatalization contrast, M1 and onset F2 are significant in predicting the posture difference in Mandarin fricatives, with both positively correlated with the palatalized category. Also by comparing the coefficients of these two parameters, onset F2 explains more variability in the model than M1. In addition, M1 and M2 interact significantly with gender. More specifically, male speakers are more likely to produce the palatalized category /ɕ/ with lower M1 and higher onset F2 values than female speakers.

To summarize, adult speakers of the three languages show categorical distinctions in different aspects of the acoustics. With both English and Japanese having a two-way contrast between a more anterior and a more posterior fricative, English speakers produce them differently mainly in the M1 dimension, whereas Japanese speakers show consistent differences between the two fricatives in M1, as well onset F2 dimensions. When Mandarin speakers produce the fricatives that contrast in place of articulation, M1 is sufficient to separate the two fricatives. For the palatalization contrast, M1 and onset F2 are significantly correlated with the tongue posture distinction, with onset F2 playing a more important role than M1 in the palatalization contrast in Mandarin.

3.5.4   Summary and discussion

To summarize, English speakers show the greatest distinctions in the M1 dimension for the /s-ʃ/ contrast, and statistical results show that M1 itself is sufficient for differentiating the two categories acoustically. This is not the case for Japanese

speakers, who distinguish /s/ from /ɕ/ in more acoustic dimensions, including M1 and onset F2. For Mandarin speakers, as for Japanese speakers, both M1 and onset F2 frequency are needed in distinguishing the three-way contrast in the sibilant fricatives.

There are complicated interactions of these three parameters together with gender, which will not be discussed further in this dissertation since the current study only includes 5 speakers per gender for each language, making it hard to generalize over larger populations. Also the focus of this dissertation is on children's speech rather than adults, and the description of adults' productions will only serve as the baseline to be compared with children. Therefore, the topic of gender differentiated speech will be not explored further in this dissertation.

The results of adult speakers' productions show that there exist crosslinguistic differences in the dimensionalities that adult speakers show contrastive differences in realizing different fricative categories. What this means for child speech development is that children must figure out the relevant dimensionalities and the relative weights in each dimension in their native language. Failure to do so may result in not being properly understood by adult listeners. The next chapter will discuss in depth about some cases that children do fail to adapt to the language-specific norms, and thus produce fricatives that adult speakers do not recognize as correct.

CHAPTER 4

CHILD ACQUISITION OF THE TWO-WAY VOICELESS SIBILANT

FRICATIVE CONTRASTS IN ENGLISH AND JAPANESE

4.1 Introduction

Both English and Japanese have a two-way contrast in voiceless sibilant fricatives. In order to first get an idea of what children's initial fricative productions are like, this chapter focuses on examining acquisition in these two languages using a small corpus of productions by children aged 2 through 3. Note that although both English and Japanese have two contrastive fricatives, the contrast is made in different articulatory aspects. As mentioned in Chapter 2, the English /s-ʃ/ contrast has more to do with the constriction place whereas the Japanese /s-ɕ/ contrast is more of a tongue posture difference. And Chapter 3 shows that the different articulatory distinctions in the two languages are also reflected in the acoustics. Where M1 as well as onset F2 are both relevant for the Japanese contrast, English adult speakers' fricative productions can be distinguished solely using M1, Therefore, it is of interest to investigate whether children start to show the language-specific contrastive patterns in their productions before age 3.

As noted in Chapter 1, the opposite stereotypical error patterns are found in these two languages. While in English, the typical error pattern is the "fronting" of /ʃ/ to a perceived [s] (Smit, 1990), Japanese-acquiring children are perceived as more

likely to substitute the more posterior [ɕ] or [tɕ] for target /s/ (Yasuda, 1970). Since the major method used in these norming studies is transcription, which describe children's productions in terms of adults' perception, the question remains as to whether the seeming difference in error patterns is the result of biases introduced from adults' language-specific perceptual norms. That is, do English-learning children and Japanese-learning children initially make the same intermediate productions which are simply interpreted differently on the way toward adult-like patterns or do English-speaking children and Japanese-speaking children behave differently as they develop their fricatives? Acoustic analysis is helpful here since it can describe children's productions without relying on adults' ears.

Another question of interest arises from the fact that the acquisition of sibilant fricatives is late compared with other consonants and usually takes a relatively longer period of time. For example, a large cross-sectional study of American English-speaking children by Smit (1990) found that at age 3, only 56 percent of children correctly produced word-initial /ʃ/ and only 62 percent correctly produced word-initial /s/. By contrast, 75 percent of English-speaking children correctly produced word-initial /f/ and over 90 percent correctly produced initial /d/ and /t/. Similarly, a cross-sectional study of Japanese 3-year-olds by Yasuda (1970) found that only 60 percent correctly produced initial /ɕ/ and only 25 percent of correctly produced /s/. Given that these contrasts undergo such a long developmental period, we may be able to find evidence even in a cross-sectional study that at least some children go through a stage of covert contrast.

Covert contrast has been defined as the production of a perceptually unreliable, but statistically significant acoustic difference between two sounds, and has been observed for a variety of contrasts, including the voicing contrast for stop consonants (Macken and Barton, 1980; Maxwell and Weismer, 1982; Scobbie et al., 2000) and stop

place of articulation (Forrest, Weismer, Hodge, Dinnsen, and Elbert, 1990; White, 2001). Many studies of covert contrast have focused on children with phonological disorders (Scobbie, 1998: for a review of different types of covert contrast), and it has been shown to be of clinical significance in that children who produce covert contrast have a better prognosis than children who produce no contrast at all (Tyler, Figurski, and Langdale, 1993).

There has been relatively little work on covert contrast in the acquisition of fricatives. Baum and McNutt (1990) observed covert contrasts in both amplitude and spectral shape between misarticulated /s/ (which was perceived as [θ]) and the target /θ/ in English-speaking children with phonological disorder. Tsurutani (2004) also found some evidence of covert contrast in the productions by younger typically developing Japanese-acquiring children of a small number of words that happened to exemplify the /s/-/ɕ/ contrast in a larger study of other contrasts. By definition, the study of covert contrast requires the use of instrumental measures in addition to a transcription analysis. For example, studies of covert contrast for stop voicing differences in English have measured voice onset time in word-initial consonant productions (Macken and Barton, 1980) and preceding vowel duration in word-final consonants (Maxwell and Weismer, 1982). In this chapter, the same five parameters that were included in Chapter 3 are used to describe children's productions.

Two predictions are made. One is that children show language-specific use of acoustic cues in fricative productions as early as their fricatives are recognized by adults as correct. The second is that covert contrast in the productions of at least some of the children will be observed, given the protracted period of acquisition of this contrast in both languages.

53

4.2 Methods

4.2.1 Participants

In order to test the above predictions, 21 child participants aged 2- year-olds and 3-year-olds were recruited, approximately ten children for each of these two age-groups for each language. All children had normal speech and language, based on parent and teacher report and had passed a hearing screening using otoacoustic emissions at 2000, 3000, 4000, and 5000 Hz. Table 4.1 gives information on the age of the child participants and the exact number of children per age group. The English-speaking children were tested in Columbus, Ohio, and the Japanese-speaking children were tested in Tokyo and Hamamatsu, Japan. All children were monolingual speakers of their native language. The children's responses were recorded directly onto a CD or a digital audiotape, using a high-quality head-mounted microphone.

Age groups:		English	Japanese
2-year-olds	age	31 (3.4)	32 (1.8)
	N	9	13
3-year-olds	ages	39 (2.6)	44 (2.4)
	N	12	9

Table 4.1: Mean age in months (standard deviation in parentheses) and number of subjects for child participant groups for English and Japanese.

4.2.2 Materials

The materials are very similar to those described in Section 3.3.2 in Chapter 3, and developed in the same way. (See Table 3.2 and Table 3.3 in Chapter 3).

4.2.3   Procedure

Children were engaged in the same word-repetition task as the adults described in Chapter 3. The task was administered by experienced speech pathologists or research assistants who usually carry out the task with typical child-direct speech register. The English-speaking children were instructed as follows: "You are going to see some pictures on my computer and hear some words. Your job is to repeat the words you hear. So if you hear the computer say 'ball', what are you going to say?" The instructions were similar in Japanese. The task was relatively simple and the children learned it easily. During the experiment, children were asked to repeat responses in the following cases: (1) if the response was different from the prompted word (e.g., the child said duck when prompted with goose) or (2) if the tester thought the target sequence would be impossible to transcribe because the response was spoken very softly, or overlapped with the prompt or with background noise (e.g., a door slam).

4.2.4   Transcription

All audible responses were transcribed and included in the statistical analysis. A native speaker/trained phonetician transcribed all initial target CV sequences, using both the audio signal and the acoustic waveform. The English data were transcribed by an American English speaker and the Japanese data were transcribed by a Japanese speaker. Both transcribers were from the same dialect regions as the child participants. The fricatives were transcribed as either correct or incorrect. The native speaker also transcribed substitution errors when the target consonant was categorized as incorrect. The transcriptions were based on the word-initial consonant-vowel sequence which the transcribers could isolate on the waveform and listen to as often as necessary. The transcribers transcribed on a child-by-child basis so that both sibilant

fricatives (and all other target word-initial obstruent sounds) were transcribed for one child before moving on to the next. The transcribers always knew what the target word (and fricative) was. A second native speaker independently transcribed 20% of the data using the same methodology. Phoneme-by-phoneme inter-rater reliability was 90% for English and 89% for Japanese.

4.3 Results

4.3.1 Transcription

Transcription analysis of elicited single word productions has traditionally been used to describe the age at which most children correctly produce a particular consonant (Smit, 1990). These "developmental norms" for consonant mastery are used primarily to help with clinical diagnosis of speech sound disorders. Mastering a consonant means that a child is able to produce the sound in a form that adult listeners accept as correct. More specifically, the operational definition in the literature for "mastery" of a speech sound typically is 75 percent accuracy for an individual child in a particular word position (Templin, 1957; Smit, 1990). Similarly, the criterion used for mastering the contrast between two sounds is 75% accuracy for both sounds in a particular word position. I adopted these operational definitions to determine how many of the English-speaking and Japanese-speaking children had mastered each of the two fricatives and the contrast between them, as shown in Table 4.2. Two observations are of interest: first, /s/ is mastered by more children than /ʃ/ in English, while more children master /ɕ/ than /s/ in Japanese; this is especially true for the two-year-old group. Further more, more English-speaking children than Japanese-speaking children have mastered /s/ by age 3. ($\chi^2(1, 22) = 6.5$, p < 0.01)

Fricative	English (N = 22)		Japanese (N = 21)	
	2-year-olds (n=9)	3-year-olds (n=13)	2-year-olds (n=12)	3-year-olds (n=9)
/s/	9	8	0	5
/ʃ/ or /ɕ/	2	6	2	5
/s/-/ʃ/ or /s/-/ɕ/ contrast	1	5	0	3

Table 4.2: Number of children with 75 percent or more correct productions in each language, based on transcription analysis.

Table 4.3 shows the most frequent substitution processes for the two languages. In English, by far the most common process is "fronting" of /ʃ/ to [s], both in terms of the number of children who produced this substitution and in terms of the total number of substitutions made by all of the children. By contrast, in Japanese, the most common process is "backing" of /s/ to [ɕ], although this substitution pattern does not outnumber other substitution processes as much as the English error pattern does. In general, fronting errors predominate in English while backing errors predominate in Japanese. In both languages, no major vowel effect on the substitution patterns was observed. In English, /ʃ/-to-[s] substitutions occurred in front of all five vowels and at a similar rate (/a/: 20; /e/: 32; /i/: 23; /o/:21; /u/:20). In Japanese, /s/-to-[ɕ] substitutions occurred in front of all four vowels where /s/ is attested - that is, before all vowels except /i/, which is a phonotactically illegal environment for /s/. Again, for Japanese, the substitution rate was similar across the rest of 4 vowels (/a/: 14; /e/: 15; /o/:16; /u/:18).

4.4   Statistical results of acoustic analyses

A set of five two-way ANOVAs were performed on each child's productions in the two languages to compare the results of the transcription analysis to the results of

	Error pattern	English			Japanese		
		Error type	Num (children)	Num (instances)	Error type	Num (children)	Num (instances)
Place error	Fronting	/ʃ/ → [s]	12	116	/ɕ/ → [s]	7	18
		/s/ →[f, v]	13	39			
		/s/ → [θ]	2	2			
	Backing	/s/ → [ʃ]	2	6	/s/ → [ɕ]	11	63
					/ɕ/ → [ç]	3	12
Manner error	Stopping	/s/ → [tʰ,t]	5	6	/s/ → [t, d]	8	25
		/ʃ/ → [tʰ]	1	1	/ɕ/ → [t,d]	7	21
		other	3	9	other	1	1
	Affrication	/s/→[tʰs, ts]	4	14	/s/→[ts, dz]	2	3
		/ʃ/ → [tʃ]	4	26	/ɕ/ → [tɕ]	7	31
		/s/ → [tʃ]	1	1	/s/ → [tɕ]	7	19
					/ʃ/ → [tʰ s]	2	2
	Other	/ʃ/ → [h]	1	1	/ɕ, s/ →[h]	2	2

Table 4.3: The most frequent substitution processes in the productions of English-speaking and Japanese-speaking children.

the acoustic analysis and also to identify instances of covert contrast. In order to describe children's productions from more acoustic aspects, in addition to the three acoustic parameters used in describing adults' productions, M3 and M4 were also included. This is because the primary goal for this chapter is to see whether children make subtle differences in the acoustic dimensions that adults may not use in their own productions. Also since each of the five acoustic parameters is entered separately into ANOVA, collinearity is not a problem for this test.

The within-subject factors were fricative and vowel, and the dependent variables were the five acoustic measures. The rows of Table 4.4 with talker ID in italics give the results of the ANOVAs for the six English-speaking and the three Japanese-speaking children who were transcribed as producing a contrast between the alveolar

and the post-alveolar fricative. It is clear from the table that for the six English-speaking children who were identified as producing clear contrasts, they all make significant distinctions in the M1 dimension, but not necessarily do so in other acoustic dimensions. By contrast, for the three Japanese children who were perceived as making clear contrasts, they all make distinctions in both M1 and M2 dimensions. This pattern suggests that adult transcribers of the two languages may use different perceptual criterion in identifying children's correct fricative productions.

Talker	Fricative spectrum				CV Transition
	M1	M2	M3	M4	OnsetF2
e2n10m	79.0 ***	27.2 ***	39.1 ***	11.9 **	16.4 ***
e3n00f	61.6 ***	6.9 *	18.8 ***	1.4	21.8 ***
e3n01m	62.3 ***	11.0 *			
e3n03f	238.9 ***	10.6 **	35.8 ***		
e3n05f	121.6 ***		7.6 *		20.2 ***
e3n11f	205.0 **		13.2 ***		12.3 **
e2n01m					8.8 *
e2n03m	11.2 *				
e3n07m				4.7 *	
e3n12m			5.1 *		
j3n01m	14.4 **	86.7 ***	28.9 ***	19.7 ***	
j3n09m	54.0 ***	75.3 ***	36.2 ***	13.9 **	176.0 ***
j3n12f	14.9 **	5.5 *			
j2n14f				9.8 *	
j3n15m					10.1 **

Table 4.4: F-values from two-way ANOVAs on all five acoustic parameters for children with contrast (italic) or covert contrast, with English-speaking children's productions in first 10 rows, and Japanese-speaking children's productions in bottom 5 rows. (***: p< 0.001; **: p< 0.01; *:p<0.05)

Table 4.4 also gives the results of the ANOVAs for the children who were consistently transcribed as producing the most typical substitution errors - i.e., [s] for /ʃ/ in English or [ɕ] for /s/ in Japanese - but who nonetheless produced a statistically significant difference in at least one dimension. These are the children whose talker

IDs are italicized. They comprised 4 of the 16 English-speaking children and 2 of the 18 Japanese-speaking children who did not have an overt contrast by the criterion of producing 75% of both fricatives correctly. These productions were classified as showing covert contrast by definition. The F-values were smaller for the children with covert contrasts, as compared to the F-values for the children with overt contrasts, suggesting that covert contrast tends to be less stable and more variable. A second finding of interest is that the covert contrasts for the productions of all but one of these children (e2n03m) were identified as significant differences in a parameter other than M1.

Examples of covert contrast are shown in Figures 4.1 and 4.2. Figure 4.1 plots the average spectra of target /s/ and /ɕ/ for two Japanese-speaking children. One of these children (j3n01m) was transcribed as having acquired the contrast between the two sibilant fricatives and his productions showed a significant difference between /s/ and /ɕ/ for all four spectral moments. For another child (j2n14f), productions of both fricatives were transcribed as [ɕ]. The acoustic analysis of her productions revealed a significant difference between the two fricatives only in kurtosis (the fourth spectral moment). The averaged spectra of the [ɕ]-for-/s/ substitutions and that of the target /ɕ/ productions are similar, except that the [ɕ]-for-/s/ substitutions have a flatter spectral peak in the spectrum than the target /ɕ/ productions, as shown in Figure 4.1.

A different pattern of covert contrast was observed for one English-speaking child (e2n01m). The productions of this child showed evidence of a contrast between the two fricatives in onset F2 frequency. Figure 4.2 plots onset F2 frequency against M1 for the productions of this child as compared to the productions of a child transcribed as having a clearly acquired contrast. It can be observed that the productions
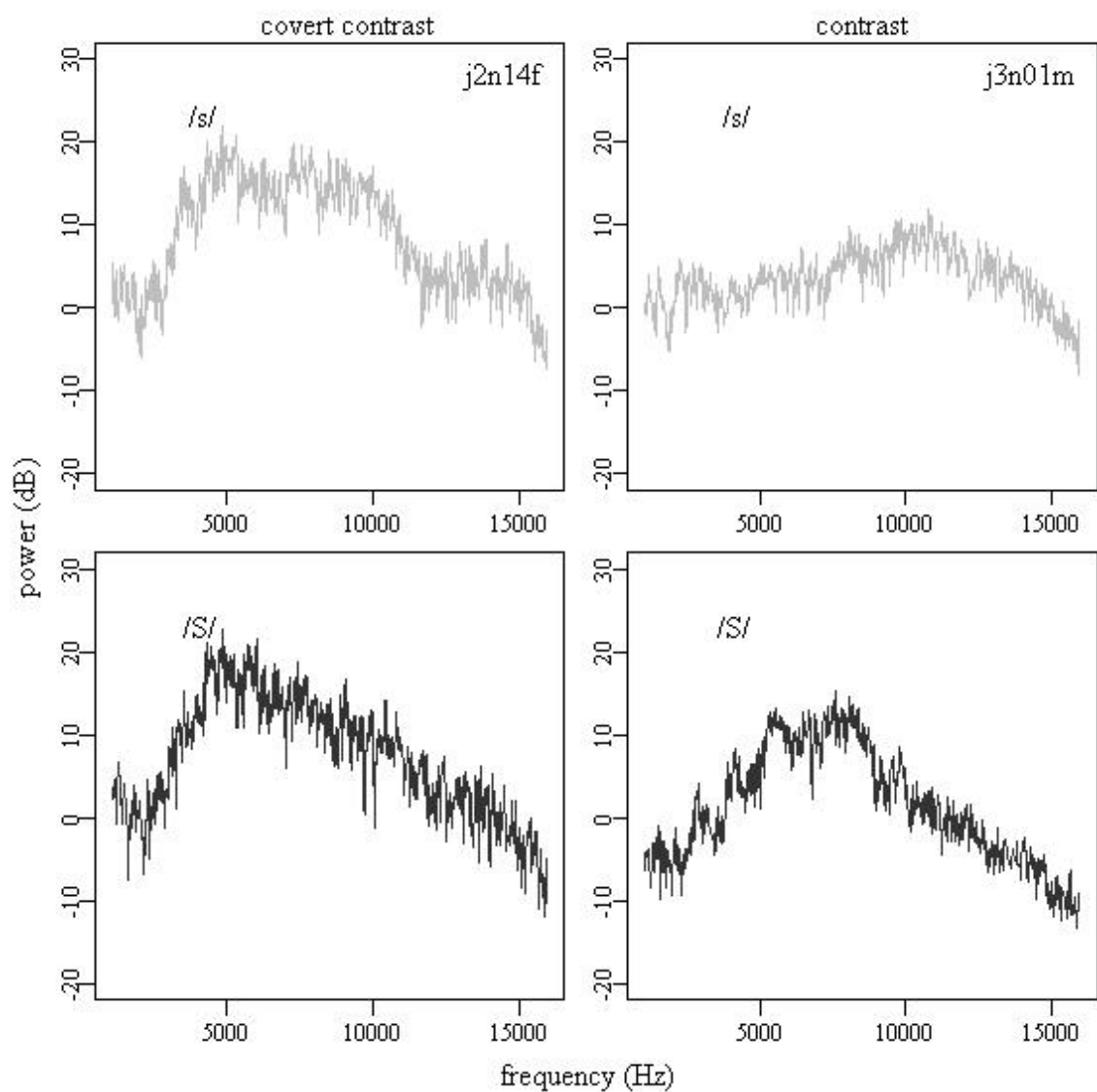
Figure 4.1: Averaged spectra of target /s/ and /ɕ/ from the productions of a Japanese-speaking child with a covert contrast (left) and a Japanese-speaking child with a clear contrast (right).

of the child with a covert contrast separate out the two categories roughly in the dimension of onset F2 frequency, with no separation in M1, whereas the productions of the child with a clear contrast make a distinction more in the dimension of M1.
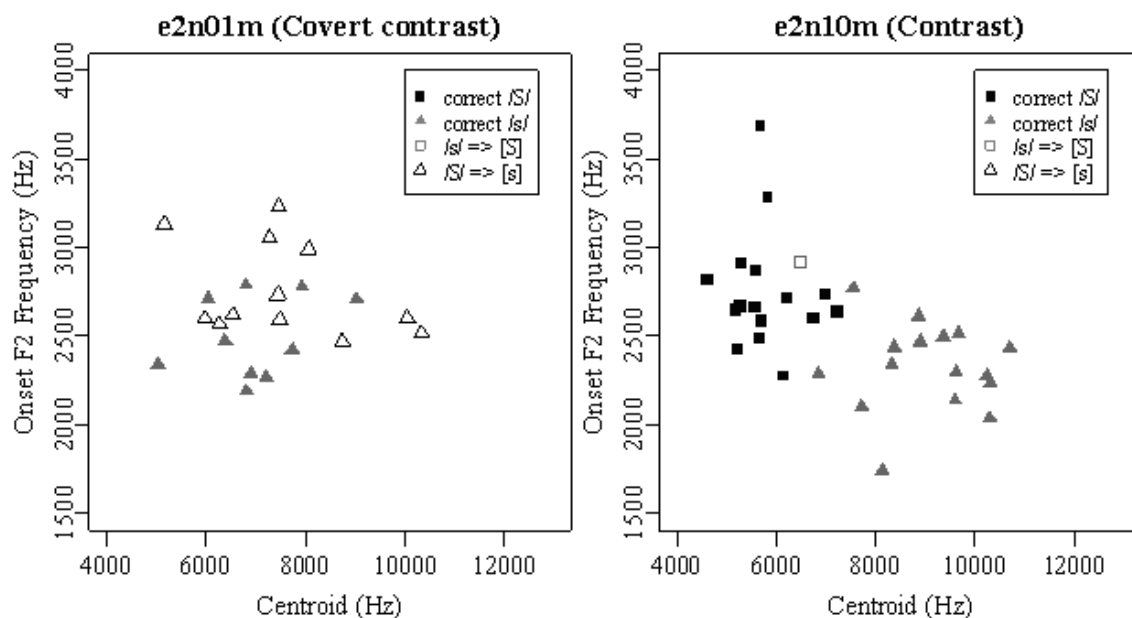
Figure 4.2: Onset F2 frequency plotted against M1 for an English-speaking child with a covert contrast (left) and an English-speaking child with a clear contrast (right).

4.5 Conclusion

Two findings are worth noting in this study. First, language-specific patterns in the productions of English-acquiring and Japanese-acquiring children were observed. Secondly, evidence of covert contrast between the two sibilant fricatives has been found in the productions of both English-speaking and Japanese-speaking children.

The cross-language differences that are clear in the children's productions involved both the transcription analyses and the acoustic analyses. More English-speaking children were transcribed as correctly producing /s/ than /ʃ/, while more

Japanese-speaking children were transcribed as correctly producing /ɕ/ than /s/. Similarly, the most common error pattern for English-speaking children was fronting ([s] for /ʃ/ and [θ] for /s/ substitutions), while the most common error pattern for Japanese-speaking children was backing ([ɕ] for /s/ and [ç] for /ɕ/ substitutions). The acoustic analyses show that children who were transcribed as producing clear contrasts use language-specific acoustic parameters that are similar to adults' patterns in their native languages. More specifically, all of six English-speaking children make distinctions in M1 dimension, whereas the three Japanese children all make distinctions in M1 and M2 dimensions.

These cross-language differences in the transcriptions might be explained in terms of the observed differences in the adult phonetic patterns in the two languages. More of the English-speaking children, as compared to the Japanese-speaking children, had acquired the contrast between the two voiceless sibilant fricatives. The later acquisition of the contrast for the Japanese-speaking children may be related to the finding in Chapter 3 that the acoustic realization of the /s-ɕ/ contrast in Japanese involves multiple acoustic parameters, as compared to the English /s-ʃ/ contrast, where a single acoustic parameter suffices to make the distinction. Moreover, the earlier-acquired contrast in English may also be related to the rounding of English /ɕ/, but not Japanese /ɕ/, which adds a visual cue for English-speaking children.

In addition, one similarity was that covert contrasts were observed in both languages. Out of 22 English-speaking children, only six showed complete mastery of the contrast at a level of 75% accuracy or more for both sounds, four children showed covert contrast as evidenced by instrumental analysis, and 12 children showed no indication of contrast mastery either in the impressionistic transcription or in the acoustic analysis. Out of 21 Japanese-speaking children, only three had mastered this contrast by age 3, two showed covert contrast, and 16 did not show any mastery

of the contrast. Two forms of covert contrast were observed. Most of the children with covert contrast used a non-primary parameter to differentiate between the two sibilant fricatives. One child with covert contrast differentiated the two fricatives with the primary parameter (M1), but the difference was not large or consistent enough to be recognized by adults. While covert contrast was observed in both languages, however, it is important to note that the direction of the emerging contrast was different. The English-acquiring children with covert contrast were just beginning to distinguish acoustically between target /s/ and transcribed [s]-for-/ʃ/ substitutions, whereas the Japanese-acquiring children with covert contrast were just beginning to distinguish acoustically between target /ɕ/ and transcribed [ɕ]-for-/s/ substitutions.

To summarize, it is found the language-specific phonetic differences affect acquisition of fricative, as judged by an experienced native-speaker transcriber. More English-speaking 2- and 3-year-old children had mastered the contrast, as compared to Japanese-speaking children of the same age. I suggest that these language-specific differences in acquisition are related to differences in how the fricative contrast is represented acoustically between English and Japanese.

Covert contrast was also observed in both languages. Four English-speaking and two Japanese-speaking children showed a significant difference between the two sibilant fricatives in one of the measured acoustic parameters in spite of the fact that the experienced native-speaker transcriber had transcribed all productions as /s/ (for English) or as /ɕ/ (for Japanese). The acoustic measures revealed cross-linguistic differences in the acquisition of the contrast as well as the presence of convert contrast. These results suggest that transcription alone is not adequate to describe phonological acquisition, since it filters children's productions through adults' perceptual norms. Acoustic analysis is a useful tool in objectively describing children's productions unbiased by adults' perception.

An interesting thing to note is that children are able to produce contrasts similar to adult production norms or different from them, with only the former being recognized by native transcribers. This seems to suggest that the acoustic parameters that adults use to categorize speech sounds are similar to the ones they use in their own productions. In other words, this seems to suggest that perception tracks production in adult speakers. Chapter 5 will systematically evaluate this possibility.

CHAPTER 5

ADULT PERCEPTION OF THE TWO-WAY FRICATIVE CONTRAST IN
ENGLISH AND JAPANESE

5.1   Introduction

Chapter 4 suggests that there is mismatch between adults' transcription and chil-
dren's production in that adult transcribers categorize children's productions using
some acoustic parameters, whereas children may make contrasts in other acoustic
dimensions. Moreover, the transcribers of the two languages seem to differ in which
acoustic dimensions they use to identify clear contrasts in children's production. More
specifically, Table 4.4 on page 59 shows that English-speaking children were identified
as making clear contrasts as long as they made distinctions in M1, but Japanese-
speaking children were judged to have a clear contrast when they made distinctions
in both M1 and M2. Since the results presented in Chapter 4 mainly reflected sin-
gle native adult's judgment for each language, it is hard to know whether different
perceptual criteria used are reflections of different perceptual norms between the two
languages or more of individual differences between transcribers. Thus the goal of this
chapter is to systematically evaluate the perceptual criteria that adult listeners use in
categorizing children's fricatives, and to find out whether there are any crosslinguistic
differences in adult perceptual norms between English and Japanese.

Previously, many studies have been conducted on the perception of voiceless sibilant fricatives in English, where the contrast is made primarily in place of articulation. These studies show that the fricative noise itself contains enough information for listeners to make identification and classification of sibilant fricatives (Hughes and Halle, 1956; Bladon, Clark, and Mickey, 1987; Evers, Reetz, and Lahiri, 1998). In particular, the spectral characteristics in the frication part are found to play a central role in perception, with /s/ typically having the peak at 4-8 kHz and /ʃ/ at 2-4 kHz for English. Meanwhile, some other studies found that formant transitions of the following vocalic context also help to cue different sibilant fricatives (Delattre, Liberman, and Cooper, 1964; Nittrouer, 1992). However, several studies suggest that spectral characteristics override contextual information such as onset F2 frequency as the primary perceptual correlate for voiceless sibilant fricatives in English (Harris, 1958; LaRiviere, 1975; Nittrouer, 1992). Transitional information, especially F2 vowel transition is relevant (Heinz and Stevens, 1961), but not crucial in fricative categorization, especially for /ʃ/ (LaRiviere, 1975; Whalen, 1984).

However, the predominant role of cuing on fricative noise instead of vowel transition is not universal for fricative perception in other languages. For example, for listeners speaking Shona, a language that makes a three-way contrast among /s/, /ʃ/ and /sʷ/, the fricative-internal cue has been found inadequate in differentiating all three fricatives, and successful identifications of the three fricatives depends on both the spectral shape of the frication noise and the transitional information. Polish is another language that makes a three-way contrast among /s/, /ɕ/ and /ʂ/, a set of fricatives that is very similar to the Mandarin set (Ladefoged and Maddieson, 1996). Nowak (2006) found that, when presented with fricative noise only, native listeners of Polish can identify the three sibilant fricatives with great accuracy, but tend to confuse /ɕ/ with /ʂ/ when fricatives are cross-spliced with vowels from other

fricative contexts, indicating an overriding effect of transitional cues in identifying the alveolopalatal fricative. The results were interpreted as indicating that adult Polish speakers apply both the fricative-internal cue and the transitional cue in classifying the three-way contrast. And more likely, the transitional cue carries more weight in differentiating the palatalization contrast.

As mentioned earlier, unlike in English, where /s/ and /ʃ/ can co-occur in front of all the vowels, Japanese /s/ never occurs before the vowel /i/ and alveolopalatal /ɕ/ can only marginally occur before the vowel /e/. This skewed distribution of vocalic contexts makes transitional cues more informative than otherwise. Onset F2 frequency in Japanese not only cues fricative articulation, it also mediates the vocalic contextual information at the same time, which constrains which fricative can co-occur. For example, Lambacher, Martens, Nelson, and Berman (2001) asked Japanese listeners to identify English /s/ and /ʃ/ contrast before different vowel contexts, and found that Japanese listeners have more difficulty in identifying /ʃ/ and /s/ next to the vowel /i/ than in other vowel contexts in word-initial and word-final position in particular. Therefore, Japanese listeners are predicted to weigh the transitional information such as onset F2 frequency more heavily than English-speaking listeners in fricative perception.

5.2   Methods

5.2.1   Stimuli

The stimuli are extracted consonant-vowel sequences from the children's word repetitions that were elicited in the production experiments described in Chapter 4, which contains productions from children aged from 2 to 3 years old. The stimuli included

68

are word-initial CV sequences both from productions of correctly-transcribed word-initial fricatives, and from productions where target /s/ was transcribed as [ɕ] (for the Japanese-speaking children) or [ʃ] for the English-speaking children and where target /ɕ/ or /ʃ/ was transcribed as [s]. The words whose initial fricatives were transcribed as having stopping errors or fricative substitution errors with /f/ or /θ/ were excluded, in order to make sure the stimuli were within the relevant acoustic dimensions that are of interest.

In addition to children' productions, the stimulus set also contained some productions from adults who were recorded as part of the method for choosing stimuli as described in Section 3.3.2. The purpose of including adult tokens is to make sure that listeners are exposed not only to immature and intermediate child productions, but also to clear adult productions so that their perceptions are not over-trained on phonetic details that are not relevant for categorical judgments. Also the perception of adult productions will serve as the baseline in evaluating listeners' perception of children's productions, if cross-linguistic differences are to be found. The reason to use recordings from adults productions elicited in developing the stimuli instead of using the adults reported in Chapter 3 is that the current dissertation is part of a larger project,and the adult recordings reported in Chapter 3 had not been collected or analyzed by the time this set of perception experiments were conducted. One problem with these adult productions, however, is that the Japanese speakers were from different dialectal regions, which introduces confounding in explaining the possible perceptual patterns which will be discussed later.

A total of 400 stimuli were selected based on the following principles: speaker language was balanced first by selecting 200 tokens from English-speaking children/adults and 200 tokens from Japanese-speaking children/adults. Within each language, children's productions were selected based on the error types described by native-speaker's

transcriptions. Specifically, for English-speaking children, 50 tokens of correct /s/ productions, 50 tokens of correct /ʃ/ productions, and 50 tokens of [s]-for-/ʃ/ substitutions were selected. Since the error patterns are extremely skewed in that there were only a few [ʃ]-for-/s/ substitutions in the database, only 8 tokens of [ʃ]-for-/s/ substitutions could be selected. The rest of the 42 English token were filled in with adult tokens. Within each 50-token set, vocalic contexts were balanced by having equal numbers of tokens for each CV sequence except for the sequence /si/ and /ʃe/, which were excluded to match the phonotactics in the Japanese fricatives (see Table 5.1 for a complete breakdown of the 400 tokens included).

The 200 Japanese tokens were selected based on the same principles, except there were 50 tokens of [ɕ]-for-/s/ substitutions, and only 11 [s]-for-/ɕ/ substitutions included because of the opposite error patterns between English- and Japanese-speaking children. Moreover, within each type, vowel context, speaker's gender and age were balanced as much as possible.

All stimuli are normalized for amplitude and ramped off to avoid unnatural sounds due to extraction.

5.2.2   Participants and task

20 English-listeners were recruited from Minneapolis, Minnesota, US, and 20 Japanese-listeners were recruited from Tokyo, Japan. The task is a speeded identification task, where each listener hears two blocks of the same 400 tokens (200 English stimuli + 200 Japanese stimuli). For English listeners, in one block, they were asked the question of whether the word initial consonant was the 's' sound. That is, the first sound in the words 'see', 'say', 'sock', 'sew' and 'Sue'. In the other block, they were ased the question of whether the word began with the "sh" sound as the first sound in the words 'she', 'shape', 'shock', 'show', and 'shoe'. For Japanese listeners, the sample

Stimulus language	Age group	Transcription type	Vowel context				
			A	E	I	O	U
English (N=200)	children	correct /s/(N=50)	13	12	0	12	13
		correct /ʃ/(N=50)	13	0	12	13	12
		[s]-for-/ʃ/(N=50)	12	0	13	13	12
		[ʃ]-for-/s/(N=8)	1	3	0	1	3
	adults	correct /s/(N=21)	6	5	0	5	5
		correct /ʃ/(N=21)	4	0	6	5	6
Japanese (N=200)	children	correct /s/(N=50)	13	12	0	12	13
		correct /ɕ/(N=50)	13	0	12	13	12
		[s]-for-/ɕ/(N=50)	12	13	0	13	11
		[ɕ]-for-/s/(N=11)	4	0	2	2	3
	adults	correct /s/(N=19)	4	5	0	5	5
		correct /ɕ/(N=20)	5	0	5	5	5

Table 5.1: The breakdown of the stimuli used in the perception experiments, as sorted by the stimulus language, the age groups of the speakers who produced the stimuli and the vocalic context the stimuli cover. Note that the sequence /si/ and /ʃe/ were not included in English in order to pair with the vocalic distribution in Japanese.

words used for the sound with the 's' label are /semi/ (*cicada*), /sori/ (*sleigh*), /sora/ (*sky*), and /su:pu/ (*soup*). The sample words used for the sound labelled as 'sh' are /ɕika/ (*deer*), /ɕeriou/ (*sheriff*), /ɕaɕin/ (*photo*), /ɕo:ri/ (*victory*), and /ɕumi/ (*hobby*).

The presentations of the two blocks were counterbalanced within the 20 English listeners as well as the 20 Japanese listeners. The order of the actual stimuli presentations inside each block was randomized for each individual listener. For each block, listeners needed to answer by pressing the 'Yes' or 'No' button as fast as they can using the index finger of their dominant hands. Naive listeners did not know that they were listening to multiple languages. Two English listeners' data turned out to be unusable and were thrown away.

The instructions for English-listeners are straightforward since the label 's' or 'sh' is transparent from the orthography. For Japanese listeners, the instructions and sample words that were used to define the 's' and 'sh' labels were written with the standard writing system, which is a mix of kanji (Chinese characters), katakana (Japanese native syllabary) and hiragana (Japanese native syllabary). Although all the sample words used contain word-initial /s/ for the 's' label or the /ɕ/ sound for the 'sh' label, these word-initial fricatives are by no means transparent or easily decomposed from the writing of the instructions.

5.3   Analysis

5.3.1   Community categorical judgments of the fricative contrast

Since the purpose of this experiment is to tap into native listener's perceptual norms, a working definition for "perceptual norm" is needed. Perceptual norm, in this study, is based on the community's opinion of correct or incorrect in judging a certain sound produced by toddlers in their native language, by pooling all subjects' responses for every stimulus token. In order to achieve this end, each token was labeled as 'correct 's'' or 'correct 'sh'' or 'neither' based on the aggregated community opinion, which was calculated as follows. In analyzing Japanese listeners' data, a token was labeled as 'correct 's'' if it receives 'Yes' responses from 14 out of 20 listeners (14 is the threshold out of 20 to be significantly different from chance, based on the binomial probability distribution) when the question being asked is 'Is this an 's'? '. Similarly, a token was labeled as 'correct 'sh'' if it received 'Yes' responses from 14 out of 20 listeners when the question being asked was 'Is this a 'sh' '. No token reached the community correctness criterion for both 's' and "sh". Those tokens which did not receive more than 14 positive responses in either block were labeled as 'neither'. In analyzing the

English data, since only 18 subjects were included, the criterion for 'correct 's" or 'correct 'sh" was to have 'Yes' responses from 13 subjects out of 18.

5.3.1.1 Community judgments on correct categories

The question of interest is how the community categorical responses would relate to the various acoustic parameters that were used to describe adults' productions in Chapter 3. Would English listeners pay attention to similar acoustic aspects as the Japanese listeners? In order to answer this question, for both listener groups, logistic regressions were performed with the dependent variable being the community categorical response (correct 's', which is 0, vs. correct 'sh', which is 1, excluding the 'neither' category which is examined later in this chapter). Because M3 and M4 again introduced collinearities and cannot be included in the final model, the independent variables were the values of the other three acoustic parameters (M1, M2 and onset F2) together with stimulus language, and the interaction between these acoustic parameters with the stimulus language. Since both English stimuli and Japanese stimuli were presented at the same time in one block to subjects without explicit instructions, including stimulus language as one of the independent variables enabled me to assess whether the subjects were aware of the mixed nature of stimuli. In cases that listeners did respond to the two languages differently, the interactions between stimulus language and each of the acoustic parameters allowed me to see whether they use any of the acoustic parameters differently in making 's'-'sh' judgment for the two languages.

Results are presented in Table 5.2, which combines the results of two logistic regressions. For the English-speaking listeners (the left half of Table 5.2), it shows that M1 (centroid frequency) and onset F2 frequency are the two parameters that primarily correlate with the community categorical judgments on the /s-ʃ/ distinction

Acoustic parameters	English listeners		Japanese listeners	
	Coefficient	p	Coefficient	p
Intercept	-1.6	**0.01**	-0.2	0.3
M1	-5.5	**<0.001**	-3.9	**<0.001**
M2	-0.6	0.2	-0.7	0.10
OnsetF2	2.1	**<0.001**	3.2	**<0.001**
Stimulus language (Japanese)	2.3	**0.003**	0.1	0.8
M1: stimulus language	0.7	0.6	0.8	0.5
M2: stimulus language	0.7	0.3	0.7	0.2
Onset F2: stimulus language	0.8	0.78	-1.2	0.2

Table 5.2: Comparison of the perceptual norms between English listeners and Japanese listeners using logistic regression, with community's categorical judgments (correct 's' vs. correct 'sh') correlating with the five acoustic parameter as well as stimulus language and the interaction between the acoustic parameters and stimulus language. Significant p values are in bold.

for both the English stimuli and the Japanese stimuli. The negative coefficient of M1 indicates that the lower M1 a fricative sound has, the more likely it is perceived by English-speaking listeners as the sound /ʃ/. This is exactly consistent with the production pattern because /ʃ/ has a lower dominant spectral energy range than /s/ and thus has a lower M1 value. Similarly, the positive coefficient of onset F2 suggests a positive correlation with the percept of the /ʃ/ category. Importantly, M1 contributes much more than onset F2 to English listeners in predicting fricative categories, as evidenced by the higher absolute value of the coefficient for M1 (*5.5*) than that for onset F2 (*2.1*). A significant effect was found in stimulus language for English-speaking listeners. Since the coefficient is positive, it indicates that English listeners are more likely to categorize a given stimulus as /ʃ/ when judging Japanese stimuli than in judging English stimuli. Meanwhile, there is no interaction between the stimulus language and any of the acoustic parameters for English listeners, suggesting

that English listeners did not use any of these acoustic parameters differently in category identification in relation to the stimulus language.

For Japanese-speaking listeners, the table (the right half of Table 5.2) shows that they also use M1 and onset F2 in differentiating /s/ from /ɕ/. The directions of these two parameters in predicting the /ɕ/ category are the same with English listener's responses. That is, M1 is negatively associated with the identification of the /ɕ/ category, whereas onset F2 is positively correlated with the /ɕ/ category. However, in terms of explanatory power, M1 (coefficient: *-3.9*) is slightly better than onset F2 (coefficient: *3.2*), which means that Japanese listeners reply almost equally on these two parameters in distinguishing /s/ from /ɕ/. Also unlike English listeners' responses, there is no significant stimulus language effect, suggesting that Japanese listeners do not treat the two languages differently in predicting fricative categories. Furthermore, there are no significant interactions between the stimulus language (Japanese) and any of the acoustic parameters, which suggests that Japanese listeners did not use the acoustic parameters differently in perceiving the fricative categories.

Since both English listeners and Japanese listeners primarily use M1 and onset F2 frequency in categorizing the contrast in voiceless sibilant fricatives, the next question is to see how this interacts with the difference in cue weighting patterns. Figure 5.1 plots the community's categorical judgments on sibilant fricatives for listeners of the two languages in relation to the two acoustic parameters, and compare their perceptions when listening to their native language with when they listen to the other language. Linear discriminant function contours were drawn to visually aid the discussion. The way that the linear discriminant function lines separate out the 's' sound from the 'sh' sound in the two-dimensional acoustic space by M1 and
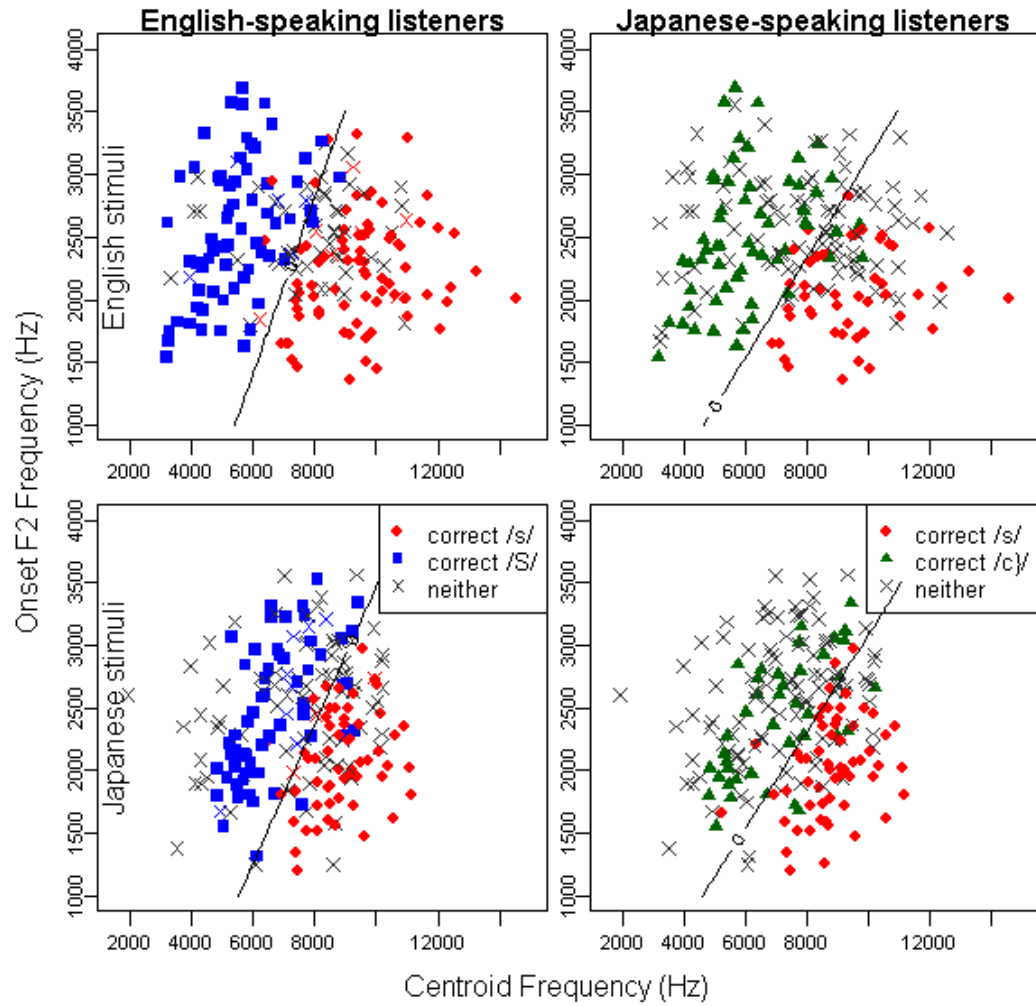
Figure 5.1: Comparison between the English-speaking listeners and the Japanese-speaking listeners in their fricative categorization patterns on the M1 and onset F2 dimensions.

onset F2 suggests the phonemic boundaries as well as the relative weights that listeners assigned in identifying fricative contrasts are differenct. More specifically, when English-speaking listeners judge their native language stimuli (the top left panel in Figure 5.1), the discriminant line roughly separates the correct 's' from correct 'sh' into left and right halves in the acoustic space. Since the x-axis of the acoustic space is the M1 dimension, this separation line indicates that M1 is primarily used, whereas onset F2 does not contribute much. Furthermore, the slope of the line does not change when English-speaking listener judge the Japanese stimuli (the lower left panel of Figure 5.1). By contrast, no matter whether Japanese listeners listen to the Japanese stimuli (the upper right panel) or the English stimuli (the upper bottom panel), the discriminant line has a shallow slope, roughly carving the acoustic space into two halves for 's' and 'sh' along the diagonal from the bottom left corner to the top right corner. Therefore, it is clear from the graph that English listeners judged their /s-ʃ/ contrast using the M1 dimension more than the onset F2 dimension. Japanese listeners' weighting strategy is different from that of English listeners in that both M1 dimension and the onset F2 dimension were used and weighed nearly equally important. Meanwhile, these language-specific weighting differences did not change when the two listener groups were asked to listen to the non-native language, as shown in the right top and the left bottom panels of the figure.

To summarize, both the statistical analysis and the graphical display of the perceptual data show that English and Japanese listeners differ in the relative weight they assign to each of the relevant acoustic dimensions in perceiving the fricative categories. English listeners mainly use M1 and onset F2 in their perception, with M1 being heavily weighted compared with onset F2. By contrast, Japanese listeners weigh M1 and onset F2 nearly equally important with their fricative perceptions.

It is also important to note that in the experimental setup with stimuli of different languages mixed together, there is a possibility that native speaker's perceptual criterion are influenced by the inclusion on nonnative stimuli, especially that English-speaking listeners show awareness of nonnative stimuli in the results. Another problem with the current experiments is that it does not ask the listeners to perceive the two sounds as a paired contrast, but rather ask them categorize individual sound. So it does not address the perceptual criteria for the fricative contrast directly. These problems will be addressed further in Chapter 7.

5.3.1.2   Community judgments of incorrect categories

It has to be noted that for both languages, there are a large number of stimuli that were judged neither as correct 's' nor as correct 'sh' by the community, called the 'neither' cases. For English-speaking listeners listening to the English stimuli, 59 stimuli out of the total of 200 fall into this case. For Japanese-speaking listeners listening to the Japanese stimuli, 103 out of 200 are such cases. Further analysis were taken to investigate the possible reasons for this large number of stimuli which were not accepted into either category.

In these analyses, the 200 tokens from each language were split into four categories, determined by the native speaker as well as nonnative speaker responses. For the English stimuli, the categories were 'correct 's'' (76 tokens) and 'correct 'sh'' (65 tokens) judged by at least 13 out of 18 English-speaking listeners. Of the 59 stimuli tokens that were not judged as either 'correct 's' or ''correct 'sh'' by 13 out of 18 English listeners, 11 were judged to be /ɕ/ by the Japanese-speaking listeners. These indicate the difference in boundaries along the relevant dimensions of the acoustic space. The rest of the 48 tokens constitute the fourth category called 'others' as they

were not identified by either listener group as one of the correct voiceless sibilant fricatives in their native language.

Table 5.3 lists the breakdown of the four categories as well as the mean and standard deviations of M1, M2 and onset F2 values of stimuli for each of the category for English stimuli. It is evident from the table that the 'neither' cases have mean M1 values (around 7900 Hz) that are intermediate between those that were judged as correct /s/ (9490 Hz) and those judged as correct /ʃ/ (5447 Hz). Moreover, the 'neither ' cases that were assimilated to /ɕ/ by Japanese listeners have very low M2 values (695 Hz) and higher onset F2 values (411 Hz) compared with those which were not (1925 Hz), reflecting a compact spectral shape and a possible palatalized tongue posture, which are characteristics of the Japanese /ɕ/ sound. The 48 stimuli that were judged in neither category by English listeners and were not assimilated into any Japanese sibilant categories have very high M2 values (1264 Hz), which suggests a rather diffuse spectral shape for these groups of sounds and may correspond to some nonsibilant fricative or fricative variant, such as /θ/ or /f/ in English.

Community's opinion		m1	m2	onsetF2
Correct	/s/(n=76)	9490 (1636)	1029 (662)	2229 (439)
	/ʃ/(n=65)	5447 (1270)	640 (493)	2568 (553)
Neither	JP /ɕ/(n=11)	7907 (1116)	695 (532)	2544 (411)
	others (n=48)	7918 (1925)	1264 (653)	2521 (333)

Table 5.3: Acoustic characteristics of the English stimuli that were judged as correct /s/ or correct /ʃ/ or neither categories by the English-speaking listeners (mean values in each cell with standard deviation in parenthesis)

Table 5.4 lists the similar breakdown of perceived categories for Japanese listeners. It should be noted that the number of 'neither' cases in Japanese stimuli is 103 out of 200, outnumbering those that were judged as either correct /s/ or correct /ɕ/. Within these 103 cases, 7 were judged by English listeners as correct /s/, 31

were judged as correct /ʃ/, and 65 were left unclassified. These unclassified tokens again have higher M2 values, suggesting diffuse spectral shapes, and are most likely to be assimilated by Japanese listeners into the nonsibilant palatal fricative /ç/ in Japanese. The very high onset F2 values for these tokens offers further support for this speculation.

Commnuity's opinion		m1	m2	onsetF2
Correct	/s/(n=57)	8838 (1128)	1225 (402)	2027 (391)
	/ç/(n=40)	7074 (1501)	1053 (416)	2407 (475)
Neither	En /s/(n=7)	9061 (798)	1297 (319)	2447 (306)
	En /ʃ/ (n=31)	6720 (1026)	1117 (393)	2591 (546)
	others (n=65)	7225 (1985)	1505 (453)	2556 (502)

Table 5.4: Acoustic characteristics of the Japanese stimuli that were judged as correct /s/ or correct /ʃ/ or neither categories by the Japanese-speaking listeners (mean values in each cell with standard deviation in parenthesis)

5.3.2 A note on listener agreement

As mentioned in Section 5.2.1, stimuli produced by both adults and children were included in the experiments. The reason to include adults' stimuli is to test how well naive listeners judge other adults' productions. By definition, adults' fricative productions should all be on target, and it is predicted that other adult listeners who speak the same language should have no difficulty in correctly identifying these productions. Figure 5.2 shows how often subjects were correct in identifying each of the adult stimuli of their native language. In each panel of the graph, the darkest bars shows the median proportion of the 18 English-speaking listeners (left) and of the 20 Japanese-speaking listeners (right) who responded "yes" to the question asking about the actual target. The top and bottom of the box shows the interquartile range and the whiskers show the full range.
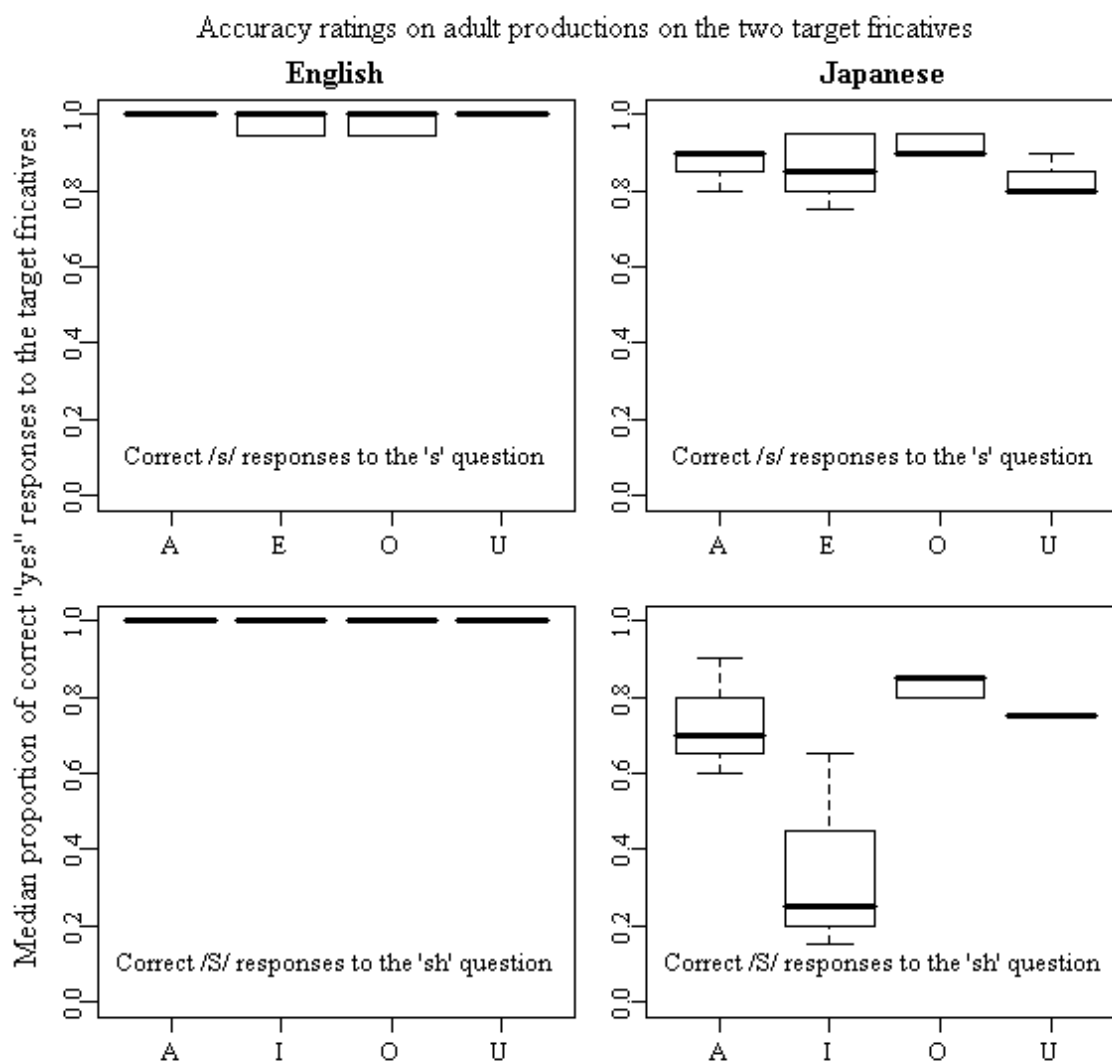
Figure 5.2: Median accuracy in judging adult speakers' productions by target and by vowel context.

It can be seen from the figure that English listeners' behavior is consistent with the prediction – they were almost 100% correct in judging the target fricatives, no matter whether it is /s/ or /ʃ/ and across the four vowels. The mean accuarcy across all vowel contexts is indicated by the horizontal line crossing all boxplots. However, Japanese listeners' reactions to the Japanese adult stimuli were very different from the predicted pattern. They showed a generally much lower accuracy in judging the target fricatives, and at the same time a bigger range of variability, which means that Japanese listeners sometimes "misinterpret" other adults' fricative productions, and not all 20 Japanese listeners agreed with each other's judgements on whether a particular token contains the /s/ target or the /ɕ/ target. Moreover, the accuracy ratings differ for different fricative-vowel sequences. For example, tokens of the /se/ sequence are judged correctly more than 80% of the time, whereas those of the /su/ sequence have a mean accuracy rate below 80%. The sequence /ɕi/ has an especially low accuracy rating of less than 30%.

There are four possible explanations for the different patterns between English listeners and Japanese listeners in judging adult's productions. One possibility is that although listeners of the two languages were given the same phoneme identification task, the metalinguistic knowledge that this task requires may not be the same between the two listener groups. This is possbily because English has an alphabetical writing system which gives its speakers a natural advantage to enhance their phoneme awareness. Japanese speakers, however, do not have the same advantage, as their writing system is syllabic (and part of Japanese writing system are logographic). The phonemic identity of fricatives is especially confusing. In the hiragana table (a table of syllabaries of consonant-vowel sequences) that Japanese speakers have to learn in their elementary school, the hiragana graphemes of /sa, ɕi, su, se, so/ are mixed in one column, giving the impression that the initial consonants in these sequences are

the same. This can possibly account for the very low accuracy rating on the sequence of /ɕi/ when they were asked the question of whether the initial consonant is 'sh' or not. (A more detailed account of the Japanese writing system and its influence on perception is offered in Chapter 7.)

The second possibility is that Japanese listeners are more sensitive to the frequencies of the CV sequences than are English listeners, as syllabary of different CV sequences is the basic unit of their writing system. Setting aside the /ɕi/ sequence, the accuracies of both targets are generally in accordance with the CV frequency in Japanese as laid out in Table 4.3 on page 58 in Chapter 2. For example, /ɕa/ is less frequent than /ɕu/, which is less frequent than /ɕo/, and their accuracy ratings follow the same order, with /ɕa/ having the lowest median rating and /ɕo/ the highest median rating.

The third possible reason is that the contrast of /s/-/ɕ/ in Japanese is not as robust as the /s/-/ʃ/ contrast in English in terms of the phonological distribution. Unlike English /s/ and /ʃ/which share the same set of vocalic contexts, Japanese /s/ and /ɕ/ sounds only contrast before the back vowels, as discussed in Chapter 2. This difference in phonological distribution could make the Japanese contrast less robust than that the English one, as reflected in the generally lower median accuracies for both targets in Japanese.

The fourth possibility is that the adult stimuli in the Japanese experiment include adults of mixed dialects. The confusion and disagreement that Tokyo listeners have may come from their unfamiliarity with the dialects in adult stimuli. This possibility will be tested in Chapter 7 by doing another experiment in which speaker dialect is controlled.

5.4 Conclusion and discussion

To sum up, both English listeners and Japanese listeners use M1 and onset F2 in differentiating the two-way fricative contrast in their native language. However, the two listener groups weigh these two acoustic cues differently in their perceptual norms, with the English listeners paying more attention to M1 than onset F2 frequency and the Japanese listeners weighing the two cues almost equally. This difference in the weighting strategy in perception is similar to adults' production patterns discussed in Chapter 3, where the English speakers produce the two fricatives differently in the M1 dimension only, whereas Japanese speakers' productions of the two sounds contrast in more dimensions.

Also, English listeners and Japanese listeners differ in how well they categorize adults' productions. English listeners are very accurate and consistent upon judging other English-speaking adults' fricatives, whereas Japanese listeners are much poorer in doing so, and much less consistent amongst themselves. This may be because of the influence of the writing system, CV frequency effect, robustness of phonological contrast, or dialectal confusion from the adults stimuli. Some of these possible accounts are tested later in Chapter 7.

CHAPTER 6

CHILDREN'S FRICATIVE ACQUISITION IN ENGLISH, JAPANESE AND
MANDARIN

6.1   Introduction

Chapter 3 and Chapter 4 described studies of English and Japanese, two languages
which have only two contrastive voiceless sibilant fricatives. Moreover, Chapter 4
examined children of age 2 or 3, using a corpus of productions in which only 6 out of
the 22 English-speaking children and 3 out of the 21 Japanese-speaking children had
already acquired the contrast. This chapter describes a large cross-sectional study
which extends the childrens' age groups to include 4 and 5 year olds and at the same
time extends the investigation to include Mandarin-speaking children.

This chapter first discusses results from transcription analyses for the three
languages. The transcription method is more elaborate than the one discussed in
Chapter 3, and has the potential to document more details in children's raw produc-
tions. Acoustic analyses are performed to examine age-related patterns in children's
fricative productions. The results of the transcription analyses and those of the acous-
tic analyses are discussed in the end, and show that the two methods can in effect
complement each other in that acoustic analyses can unveil the low-level developmen-
tal patterns that are not captured by transcription results, but transcription results
do provide interpretations for the direction of the categorical split.

6.2   The corpus of productions

For each of the three languages examined, the participants were 100 children aged 2 through 5, with 25 children per age group. For the purpose of this dissertation, 40 children transcribed by native speakers in time for use (10 per age group) were included in the analysis. These participants were engaged in the identical task using the same materials as the 30 adults in the production experiments reported in Chapter 3. Like the adults, they were recorded in their home country (i.e. Columbus, Ohio for the English-speaking children, Tokyo for the Japanese-speaking children, and Songyuan for the Mandarin-speaking children). None of the children participants overlapped with those 2-year olds and 3-year olds examined in the production experiments reported in Chapter 4.

6.3   Transcription analyses

6.3.1   Method

For all forty children examined for each language, all the target fricatives were transcribed by an experienced native speaker/phonetician using the transcription conventions developed in paidologos project (http://www.ling.ohio-state.edu/ẽdwards/). More specifically, the transcriber was asked to make both a broad transcription and a narrow transcription.

The broad transcription is essentially accuracy rating, where the native speaker judged whether the initial consonant was produced correctly or not. During this process, the transcriber listened to the word production, and compares the word-initial consonant and the following vowel to their target transcriptions. If a sound is judged to be correct, it was given the label '1', and a sound judged as incorrect was

given '0'. For the ones that were perevied as off target, the transcriber tried to classfiy the errors into the three substitution categories listed in Table 6.1. More specifically, the symbol '$' is used when the substituted category is a sound within the native phonological inventory of the language. The symbol ':' indicates an intermediate or ambiguous quality to the sound being perceived, and connects the two sounds between which native speaker's perception alternates. The order of the two sound connected by the symbol indicates the listener's inclination or preference between the two, with the one to the left of the ':' sign is the primary percept, whereas the one on the right is the alternative percept. The '+' sign means a substitution with a nonnative sound, or a sound that is *outside* of the sound inventory of the native language. Taking English as an example. For a word that starts with the sound /s/, if the transcriber transcribed it as '$t^h', it means that she perceives a substitution for the target using [t^h], and at the same time the [t^h] sound is within the phonological inventory of English. '$t^h:+t$\varepsilon^h' indicates a substitution error where the perceived sound is ambiguous, intermediate between the [t^h] sound that English has in its inventory and the [t$\varepsilon^h] sound that is outside of the inventory of English. Moreover, the perceived sound is more like the one to the left of the ':' sign, which is the [t^h] sound, than the one on the right, which is the [+t$\varepsilon^h] sound. Table 6.1 summarizes the transcription conventions and provide some examples as well.

6.3.2 Results

Figure 6.1 shows the results of the broad transcriptions. The accuracy percentage was calcuated by diving the tokens judged to be correct (tagged as '1') are calculated by the total number of productions in each of the 2 or 3 target sounds for each of the four age groups for the three languages. In English, except the the youngest children (age 2), the accuracy for children's /s/ is lower than that for /ʃ/. In Japanese, the

87

Transcription symbol	meaning	examples
$	substitution with a native category	t^h^: substituted by /t^h/, which is a native sound in English
+	substitution with a nonnative category	+tɕ^h: substituted by /tɕ^h/, which is not in the English phonological inventory
:	intermediate between two categories. Left to the colon is the primary percept.	t^h^:+tɕ^h: Intermediate between a /t^h/ substitution and a /tɕ^h/ substitution, but it sounds more like a /t^h/ than a /tɕ^h/ sound.

Table 6.1: Summary of the narrow transcription convention used in the native speaker transcriptions

accuracy of /s/ productions is consistently lower than /ɕ/ for all four age groups. In Mandarin Chinese, the accuracy for /s/ is the lowest, and that for /ɕ/ is the highest across all four groups.

One thing to note is that the results for the English-speaking children are different from the transcription results presented in Chapter 4 for a different set of English-speaking 2 and 3 year olds. In Chapter 4, the transcription results showed that there were more children who had acquired /s/ than those who acquired /ʃ/. The difference may come from the different transcription methods that were used. The transcription method reported in Chapter 4 was not as fine-grained in the sense that there was no provision for identifying intermediate sounds. This seems to have encouraged a more categorical inclusion judgments for "seemingly" correct /s/ tokens. Similarly conflicting results are found in comparing previous studies. Templin (1957) used broad transcription and found that /s/ was acquired earlier than /ʃ/, whereas Ingram (1981) used more narrow transcription analysis, and found the opposite pattern.

This might be because children's initial /s/ productions are not adult-like, but could fit into the /s/ category if native transcribers were not very strict in making the categorization. However, if narrow transcriptions were applied so that these productions were described with more detail or using intermediate and nonnative categories, native transcribers may not count them as correct 's' as with broad transcription. This, again, suggests that the perceived acquisition orders reported in the literature may be related more with how adults were asked categorize children's productions than how children actually produce these fricatives.
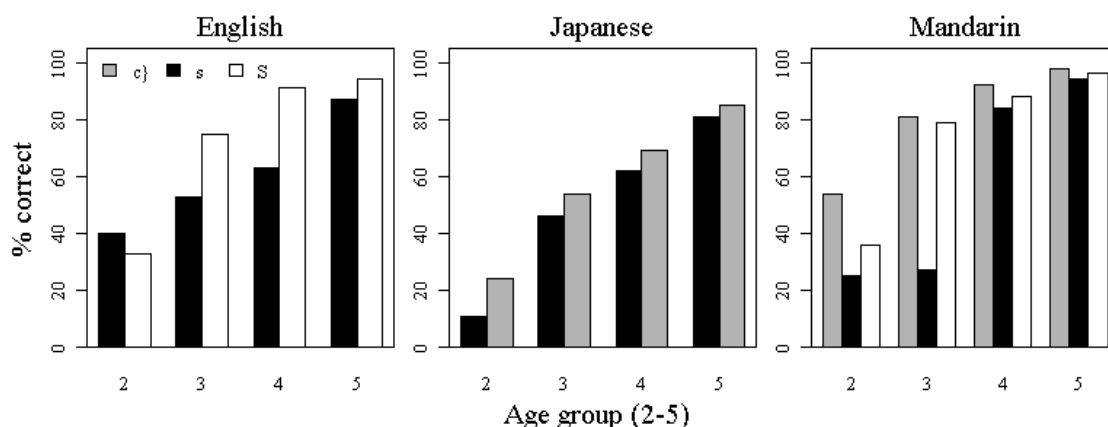
Figure 6.1: Comparison of accuracy ratings for sibilant fricatives across the three languages

Table 6.2 lists the substitution patterns in English for the two target sibilant fricatives produced by 40 English-speaking children. The table includes the broad sound categories that were used to substituting the target fricatives as well as the specific error patterns described using intermediate sound categories. It is clear from the table that the most frequent substitution process for the sound /s/ is the [θ]-for-/s/ substitution. However, among the total 48 [θ]-for-/s/ substitution cases, 21 fall into the prototypical [θ] region for the native speaker transcriber, whereas 21 are

intermediate between /θ/ and /s/. This shows that a large proportion of the [θ]-for-/s/ substitution prevalent in the studies of English children's fricative acquisition are intermediate tokens that may not be adult-like /θ/. Sarah K. Schellinger, who is the native speaker transcriber for the current set of English-speaking children, actually looked into these intermediate categories between [s] and [θ] in her Master thesis (Schellinger, 2008). She asked a large number of English-speaking listeners to judge the tokens of 'correct /s/', '$θ:s', 's:$θ', and 'correct /θ/' according to her transcription, and found that the reaction time and the correct identification rate for the intermediate categories are different from those for either the 'correct /s/' and 'correct /θ/' tokens, and therefore provide empirical evidence for listener's awareness of such intermediate categories.

Similarly, when English-speaking children produce the /ʃ/ sound, the [s] substitution is the most frequent substitution process. However, 9 out of 39 [s]-for/ʃ/ substitutions are intermediate cases, which again shows that these classical [s]-for/ʃ/ substitutions may not necessarily be adult-like [s] sounds.

For the Japanese-speaking children, as evident in Table 6.3, the sound /s/ is most commonly substituted with the [ɕ] sound, whereas the most common substitution for the sound /ɕ/ is with [tɕ] or [s] sound. Similar with the English results, Japanese-speaking children produce some intermediate productions, which were captured by the error patterns using narrow transcriptions. For example, out of 32 [s]-for/ɕ/ substitutions, 6 are intermediate between /s/ and /ɕ/. One thing to note is that [t]-substitutions for target /s/ is the second most frequent error type, and frequent stopping errors for Japanese-speaking children are somewhat unique as they do not occur in English-speaking children or Mandarin-speaking children.

Table 6.4 presents the error patterns for the Mandarin-speaking children. It is shown that Mandarin-speaking children tended to produce some fricative-like sounds

Target fricative	Substitutions	Proportions	Transcribed errors	Proportions
/s/	θ	0.23 (N=48)	$θ	0.10(N=21)
			$θ:s	0.10((N=21)
			$θ:+tʰ	0.02 (N=4)
	ʃ	0.20 (N=41)	$ʃ	0.16 (N=33)
			$ʃ:s	0.02 (N=4)
			$ʃ:#	0.01 (N=2)
	tsʰ	0.16 (N=33)	+tsʰ	0.07 (N=13)
			+tsʰ:s	0.05 (N=11)
			+tsʰ:$θ	0.01 (N=3)
			+tsʰ:+tʃʰ	0.01 (N=2)
/ʃ/	s	0.30 (N=39)	$s	0.19 (N=24)
			$s:ʃ	0.07 (N=9)
			+sθ	0.02 (N=2)
	tʃʰ	0.22 (N=28)	$tʃʰ	0.15 (N=19)
			$tʃʰ :ʃ	0.02 (N=3)
			$tʃʰ :+tsʰ	0.02 (N=2)
	tsʰ	0.22 (N=11)	+tsʰ	0.05 (N=7)
			+tsʰ :$tʃʰ	0.02 (N=2)
			+tsʰ :ʃ	0.02 (N=2)

Table 6.2: Substitution patterns for English fricatives as transcribed by a native speaker phonetician.

when making substitutions for the three target sibilants, unlike English- or Japanese-speaking children, who more commonly made affrication errors or stopping errors than fricative errors. Moreover, the most frequent substitutions happen among the three sibilant fricatives, where [ʂ] and [ɕ] were used to substitute for target /s/, [ʂ] was used to substitute for /ɕ/, and [s] was most often used to substitute for /ʂ/.

Target fricative	Substitutions	Proportions	Transcribed errors	Proportions
/s/	ɕ	0.23 (N=55)	$ɕ	0.19(N=47)
			$ɕ:s	0.03((N=8)
	t	0.14 (N=33)	$t	0.12 (N=28)
			$t :+tj	0.02 (N=3)
	tɕ	0.12 (N=30)	$tɕ	0.07 (N=18)
			$tɕ:+tj	0.02 (N=6)
			$tɕ:$ts	0.02 (N=3)
/ɕ/	tɕ	0.22 (N=46)	tɕ	0.21 (N=42)
			tɕ: $kj	0.01 (N=2)
	s	0.16 (N=32)	$s	0.09 (N=19)
			$s :ɕ	0.03 (N=6)
	ç	0.13 (N=26)	$ç	0.09 (N=19)
			$ç:ɕ	0.03 (N=7)

Table 6.3: Substitution patterns for Japanese fricatives as transcribed by a native speaker phonetician.

6.4   Acoustic analyses

6.4.1   Overview

Transcription analyses are more or less constrained by native transcriber's categorical judgment which has been shown in Chapter 5 to be language-specific. Even with a transcription method that encourages categorizations using nonnative and intermediate sounds, the native language-specific perceptual criteria are not likely to be overridden completely. Therefore, although narrower transcriptions can be informative for understanding the developmental norms within each particular language community, it is hard to compare the developmental patterns across languages. Acoustic analysis can circumvent this problem by being able to describe children's fricative productions using the same set of acoustic parameters, to compare the raw patterns without them being filtered through adults' ears.

Target fricative	Substitutions	Proportions	Transcribed errors	Proportions
/s/	ʂ	0.47 (N=125)	\$ʂ	0.34(N=91)
			\$ʂ:s	0.04((N=12)
			\$ʂ:\$ɕ	0.02((N=6)
			+ʂr	0.02((N=6)
			\$ʂ:+h	0.01((N=4)
	ɕ	0.13 (N=35)	\$ɕ	0.10 (N=26)
			\$ɕ:\$ʂ	0.01 (N=4)
			\$ɕ:+h	0.01 (N=2)
	θ	0.11 (N=29)	+θ	0.07 (N=20)
			+θ:\$ʂ	0.01 (N=2)
			+θ:s	0.01 (N=2)
/ɕ/	ʂ	0.50 (N=59)	\$ʂ	0.28 (N=33)
			\$ʂ:+h	0.17 (N=20)
			\$ʂ: \$s	0.02 (N=2)
	ç	0.09 (N=11)	+ç	0.09 (N=11)
	tɕʰ	0.08 (N=8)	\$tɕʰ	0.05 (N=6)
			\$tɕʰ :ɕ	0.02 (N=2)
/ʂ/	s	0.24 (N=37)	\$s	0.16 (N=25)
			\$s :ʂ	0.06 (N=9)
	ɕ	0.20 (N=30)	\$ɕ	0.07 (N=11)
			\$ɕ:ʂ	0.07 (N=11)
			\$ɕ:\$s	0.02 (N=3)
			\$ɕ:+h	0.01 (N=2)
	θ	0.08 (N=13)	+θ	0.03 (N=6)
			+θ:\$s	0.01 (N=2)
			+θ:ʂ	0.01 (N=2)

Table 6.4: Substitution patterns for Mandarin fricatives as transcribed by a native speaker phonetician.

The same procedures used in Chapter 3 for the adult productions applied to the children's productions here. Two predictions were made. First, children, no matter what language they speak, all start from an undifferentiated gesture in their early fricative productions, and then separate their productions into different contrastive categories as their age increases. Second, the differentiation of the contrastive categories only emerges in the relevant acoustic dimensions that adult use in making the contrast. In other words, in this process of category differentiation or category acquisition, children attend to those language-specific acoustic aspects that are useful in making the contrasts in their native language.

6.4.2   Averaged Spectra in different age groups

In order to see how children start to form different fricative categories and their fricative development over age, their raw fricative spectra were averaged in different age groups to see the general tendencies of category emerging in the fricative spectrum. The averaging procedure is the same as the adults' averaged spectra which is described in section 3.4.1 on page 35. All children's fricative productions were included in the averaging process, regardless of whether the productions were correct or not by native transcribers. For the incorrect ones, which include substitutions of affricates, stops and fricatives, some substitutions using voiceless sibilant fricatives were included. Further, the raw fricative productions were averaged in terms of the canonical fricative targets, which are assumed to be the intended targets that children aimed at.

Figure 6.2 plots the mean spectra averaged over all fricatives produced for each target type by children in each age group for children speaking each of the three languages respectively. The target fricative spectra were color coded as follows: black
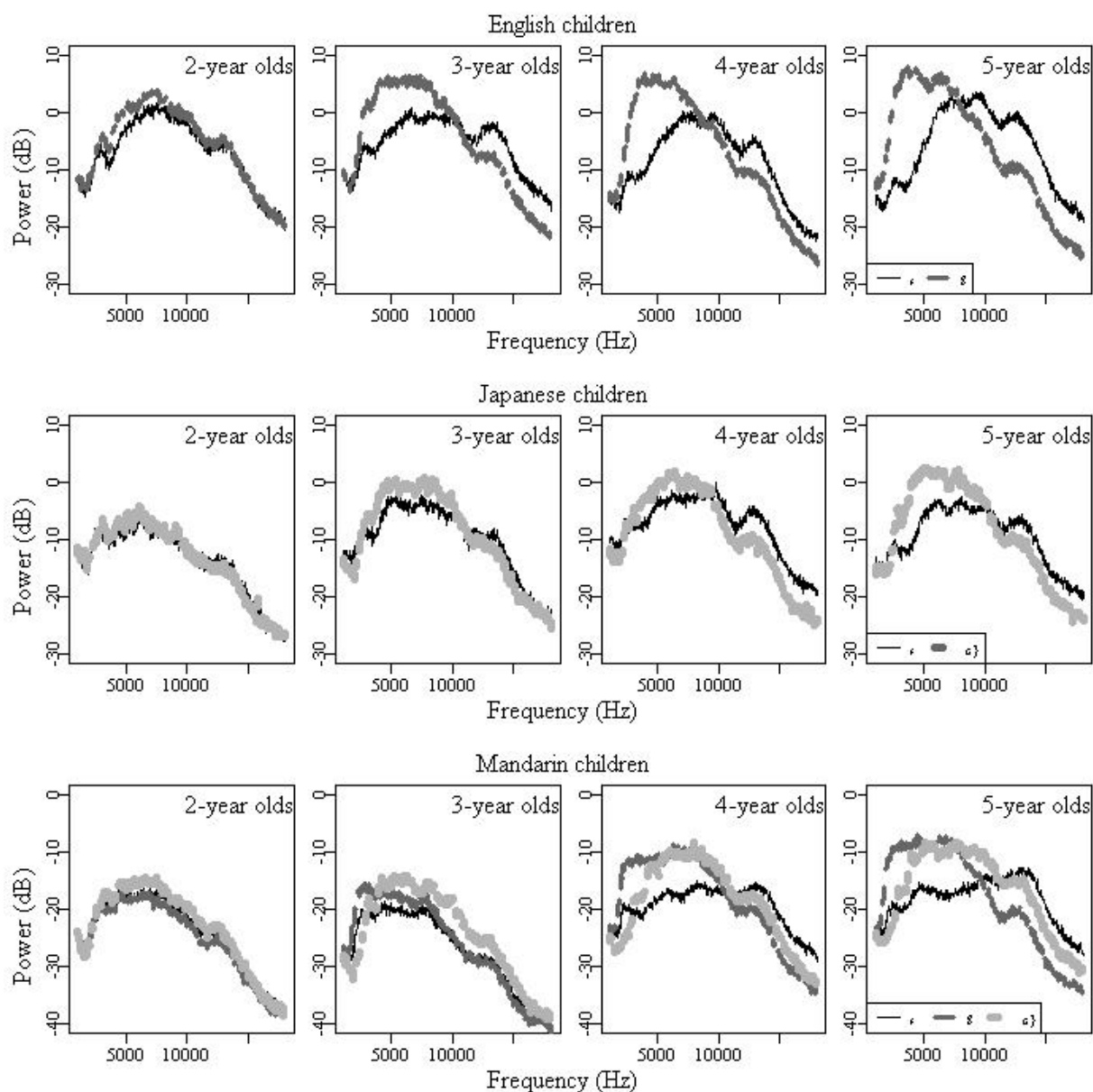
Figure 6.2: Averaged fricative spectra in different age groups for English-, Japanese-and Mandarin-speaking children. Black lines for the alveolar/dental /s/, gray lines for /ʃ/, and light gray for alveolo-palatal /ɕ/.

for alveolar/dental /s/, gray for postalveolar /ʃ/ in English as well as the retroflex /ʂ/ in Mandarin, and light gray for alveolo-palatal /ɕ/ in Japanese and Mandarin.

For English-speaking children (the upper row of panels), at age 2, the averaged spectra of the two fricative categories are almost on top of each other, and are not clearly separated from each other at all. An /s-ʃ/ distinction starts to emerge at age 3, where the spectrum of /s/ is flatter and has a peak more toward the higher frequency range. By contrast, the spectrum of /ʃ/ is peakier, and has a peak more toward the relatively lower frequency range. In age 4 and 5, the distinction is further evident – the peaks of both fricatives are more sharpened, and are further separated from each other. It is also clear from the graph that the initial undifferentiated fricative productions at age 2 do not resemble any of the target fricative shapes that emerged later. Also the maturely developed fricative spectra at age 5 resemble the English-speaking adults' averaged fricative spectra in Figure 3.3 on page 37.

Similarly, for Japanese-speaking children (the middle row of panels), at age 2, the averaged spectra of /s/ and /ɕ/ are completely overlapped, indicating a non-differentiated lingual gesture. Over the age groups, the two fricatives show gradual differentiation. At age 5, the two are distinct from each other, with the /s/-spectrum being less peaky than the /ɕ/ spectrum and the peaks of the two fricative spectra being different from each other as well. Compared with the English-speaking children's pattern, both similarities and differences can be found. What is similar between the two languages is that, for both languages, the two fricative categories start from one unimodal distribution, where no clear distinctions can be observed. They also both show category emergence from the initial undifferentiated entities to distinct patterns that resemble adults' norms. The difference seems to exist in how quickly the two categories emerge from the initial state. For English-speaking children, by age 3, the distinction is already clear, whereas that for Japanese children is less so.

In addition, the initial undifferentiated spectra for the two languages are slightly different – English-speaking children start from a peakier spectral shape and in a higher frequency range at age 2 compared with their Japanese peers.

The pattern for the Mandarin-speaking children is similar to those of the other two languages in that an undifferentiated unimodal distribution is found at age 2, and then the three categories start to separate out from each other at later ages. By age 5, all three categories have distinctive patterns that are similar to the Mandarin-speaking adults' norms. That is, the /s/-spectrum has the highest-frequency peak among the three, the /ʂ/-spectrum is lowest, and that of /ɕ/ is in between /s/ and /ʂ/. It should be noted that the 5-year old /s/ spectrum is less peaky than that for the other two fricatives. This is different from the Mandarin adults' /s/ spectrum on page 37, which is similar in peak amplitude to /ʂ/ and /ɕ/. This may suggest that Mandarin children have not completely mastered an adult-like /s/ by age 5.

6.4.3   Relationship between fricative development and acoustic dimensions

In order to see the relationship between category emergence and the different acoustic dimensions that adults in different languages use, the fricative productions plotted in Figure 6.2 were further analyzed using the three measures that were predictive of adult categories in the logistic regression analyses described Chapter 3.

Figure 6.3 plots the category emergence pattern for each language in each of these three acoustic dimension: M1, M2 and onset F2 frequency. More specifically, for each parameter, the acoustic values of all the tokens for each fricative targets produced by every individual child were plotted against their age (in months) to show the age-related change in each of the acoustic dimensions. For each target fricative, a regression line was calculated for all the productions by children, with the dependent variable being the acoustic values of each of the acoustic parameters, M1,
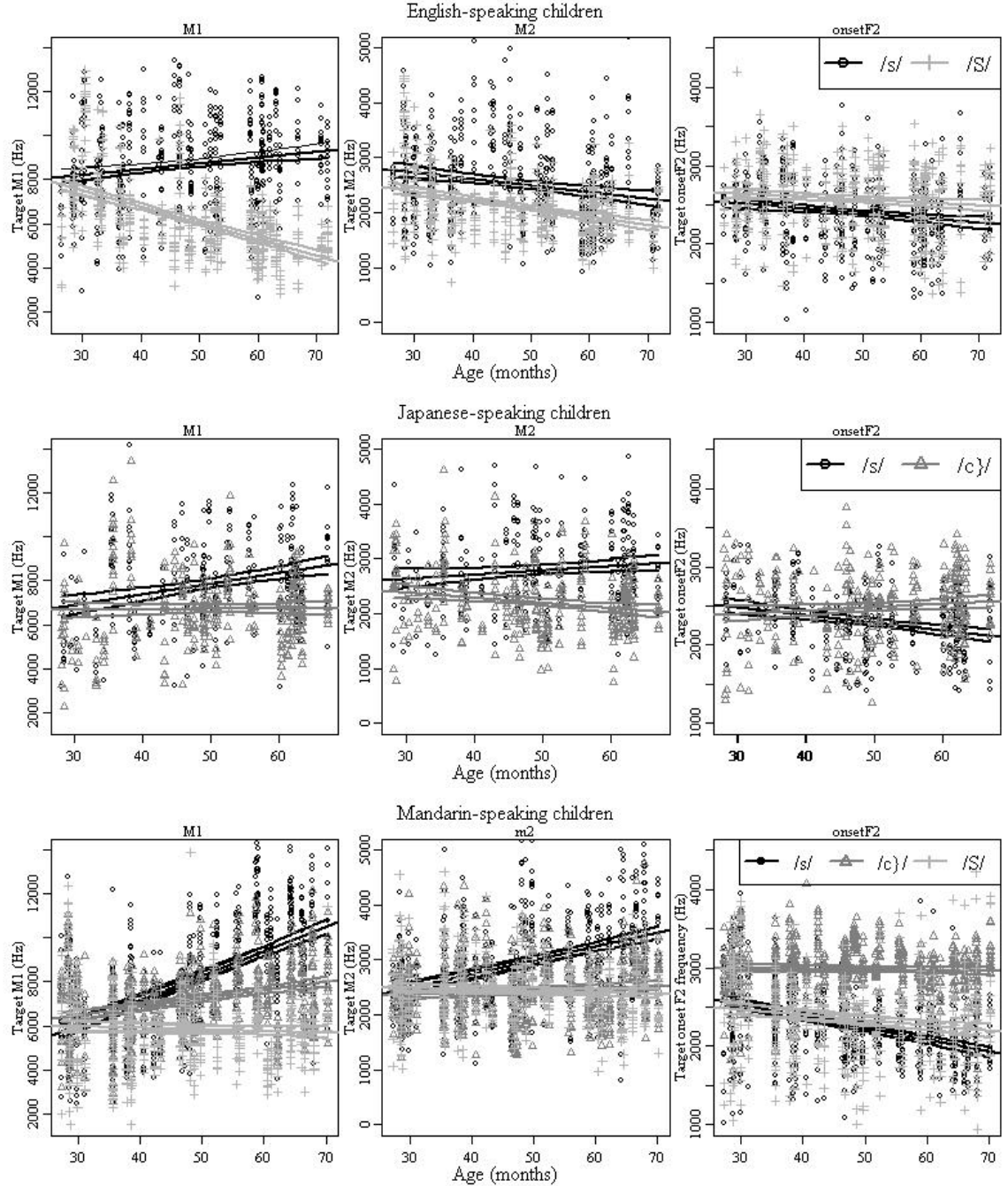
Figure 6.3: Comparison of categorical split for children of the three languages in the three acoustic dimensions as a function of age. Regression lines were fit on the acoustic values for each target fricatives over all the productions of each child against child age in different acoustic dimensions. Above and below each regression line, the 95% confidence intervals for the regressions were also drawn. Black lines and symbols represent /s/ in all three languages, grey lines and symbols are for /ʃ/ in English and /ʂ/ in Mandarin, and silver lines and symbols are for /ɕ/ in Japanese and Mandarin.

98

M2 or onset F2, and the independent variable being children's age in months. In addition, 95% of the confidence interval was also calculated for the regression line of each target in each of the acoustic dimension, to quantify when the regression lines for two/three fricative targets show significant divergence in regression slopes.

For the English-speaking children (the upper row of panels), in the M1 dimension, it is shown that the regression lines for the two target fricatives, /s/ and /ʃ/, started to show differentiated slopes as early as 30 months , with no overlap in the 95% of the confidence interval bands between the two target fricatives. And the two regression lines diverge further as children's age increases. The pattern in the other two dimensions, M2 and onset F2, are different, however. While in the M1 dimension, there is a clear tendency of category splitting as age increases, the two targets remain by and large parallel in the dimension of M2 and overlap greatly in the dimension of onset F2.

By contrast, the Japanese-speaking children show a different developmental pattern in the dimensionalities that the two fricative categories were differentiated (middle row of Figure 6.3). In the M1 dimension, they showed systematic category split around 40 months, later than the English-speaking children. Moreover, the amount of the divergence is much less than that in English-speaking children. Japanese-speaking children also show a similar amount of divergence in the M2 dimension around 30 months and a smaller amount of divergence in the onset F2 dimension around 45 months.

More complicated patterns are shown in Mandarin-speaking children, who must develop a three-way contrast in the voiceless sibilant fricatives (bottom row of Figure 6.3). In the M1 dimension, all three sounds show category split. However, the split seems to happen between /ʂ/ and the other two sounds first around 30 months. Then, a second split happens around 40 months between the other two sounds, /s/

and /ɕ/. In the M2 dimension, /s/ was separated from the other two around 30 months, whereas the other two sounds have overlapping confidence intervals which make them undifferentiated, suggesting no age-related change in these two categories in the M2 dimension. In the onset F2 dimension, a separation between the sound /ɕ/ and the other two sounds are well demonstrated before 30 months, but there is no age-related change in /s/ and /ɕ/ in this dimension. In short, all three acoustic dimensions show age-related differentiations for the Mandarin-speaking children.

To summarize, English-speaking children show category differentiation in M1 only, whereas Japanese children show a comparable split in both M1 and M2 dimensions, as well as a smaller degree of differentiation in the onset F2 dimension. In the M1 dimension, the magnitude of divergence between /s/ and /ʃ/ in English is much bigger than that between /s/ and /ɕ/ in Japanese. Mandarin children show an emerging differentiation for all three categories in the M1 dimension, the split between /s/ and /ɕ, ʂ/ in the M2 dimension, and the differentiation between /ɕ/ and /s, ʂ/ in the onset F2 dimension. Importantly, for each language, the patterns of which acoustic dimensions are targeted for category differentiation and the degree of separation in the targeted dimensions coincide with the adults' production patterns. Specifically, English-speaking adults show a systematic difference between the two categories in the language only in the M1 dimension, and this is the only dimension in which English-speaking children show a clear category divergence. By contrast, Japanese adults make fricative distinctions in the dimensions of M1, onset F2, and gender related differentiation in M2, and Japanese-speaking children's productions also show a divergence between /s/ and /ɕ/ in all three dimensions. Also, adult English speakers show a bigger M1 difference between their two fricatives than do adult

Japanese speakers, and in the same way, English-speaking children show a larger difference between the endpoint values for the two regression lines in the M1 dimension than do the Japanese-speaking children.

6.5   Discussion and conclusion

One thing that is common across the three languages is that children show an undifferentiated unimodal distributions in their fricative productions, no matter how many contrastive categories they have in the ambient languages. They all show age-related category differentiations in their productions, but only in the acoustic dimensions that are relevant for their native language. For English-speaking children, differentiation is evident most clearly in the M1 dimension, which is the primary acoustic correlate that adult speakers of English used in their productions in Chapter 3, as well as in their perception of children's productions in Chapter 4. Japanese children, by contrast, show category differentiation in the M1, the M2 and the onset F2 dimensions, all three of which are also important for the /s-ɕ/ contrast in Japanese adults' production. And Mandarin-speaking children, similar to Japanese children, show differentiated developmental patterns in all three acoustic dimensions.

At the same time, however, depending on the specific language that children speak, there are some differences as well. Firstly, where the intercept point for the initial undifferentiated pattern in each dimension is different from language to language. Secondly, the magnitude of final category difference in each dimension is different for each language. In English, the category differentiation is most obvious in the M1 dimension, whereas the differences in the final intercept values for the two curves in the M2 and onset F2 dimensions are minimal. In Japanese, by contrast, a clear category separation is shown in all three dimensions, but the degree of differentiation in the M1 dimension is much less than that in English-speaking children. For Mandarin

children, all three categories show clear differentiations in the M1 dimension, but only /s/ is separated from /ʂ/ and /ɕ/ in the M2 dimension and /ɕ/ is separated from /s/ and /ʂ/ in the onset F2 dimension.

Another thing that is shown in Figure 6.3 is that the slopes of each regression line are different for different fricative categories in the three languages. These slopes can be used to predict the relative sequence of fricative acquisition, since the steeper a slope is, the more deviant the final target realization is from the original undifferentiated gesture, and therefore the more likely the sound is to be acquired relatively late. In the M1 graph for English, both targets have relatively steep slopes, but the slope for /s/ is less steep than the slope for /ʃ/, indicating an earlier acquisition of /s/ than of /ʃ/. In Japanese, /ɕ/ has a more gradual slope than /s/ in all three acoustic dimensions, suggesting that /ɕ/ is likely to be the one acquired first. In Mandarin, /s/ has a steeper slope than the other two fricatives in all three dimensions, which is likely to make it the last sound to be acquired. It is hard to predict the relative sequence between /ʂ/ and /ɕ/, since /ɕ/ has a less steep slope than /ʂ/ in the M1 dimension whereas it is the opposite in the onset F2 dimension.

Both approaches, transcription analysis and acoustic analysis, provide converging evidences for similar fricative acquisition patterns in each language. For example, in English, children start from some intermediate fricative-like nondifferentiated gesture, then separate out the two categories. Moreover, the initial state is closer to /s/ than to /ʃ/, which is why /s/ has been claimed to be acquired earlier than /ʃ/ in transcription studies, especially when broad transcription was used. In Japanese, /ɕ/ is acquired earlier than /s/, which is confirmed by both types of analyses. In Mandarin, /ʂ/ and /ɕ/ precede /s/, as shown in both the transcription analyses and the acoustic analyses.

Chapter 5 shows that adult perceptions conform with adult production patterns for English and Japanese, two languages that have a two-way contrast, and are similar in children's production patterns as well. It is of interest to extend the perception test to Mandarin, alanguage that has a three-way sibilant fricative contrast, and see whether similar relationships can be found. Such perception test across all three languages are discussed in the next chapter.

CHAPTER 7

ADULT PERCEPTIONS IN ALL THREE LANGUAGES

7.1 Introduction

This chapter describes a new set of perception experiments. These experiments differ from Chapter 5 in four important aspects. First, they include Mandarin-speaking participants in addition to English- and Japanese-speaking ones so that a full set of crosslinguistic comparisons on the three languages can be made on adults' perceptual norms. Second, they include stimuli from all three languages. Third, the adult stimuli in the Japanese experiments this time only include Tokyo speakers in order to test whether the lower accuracy rates found in Japanese listeners in the previous experiments in Chapter 5 came from dialectal differences. Fourth, these experiments include stimuli produced by children of age 2 to 5, so that adults' perceptions are not too over trained by younger children's productions to affect their perception criterion.

The primary objective of this set of experiments is to expand the crosslinguistic comparison to adult perception patterns in all three languages. The results of the perception experiments reported in Chapter 5 showed that adult perceptual norms track production norms for English and Japanese speakers. The current experiments are designed to test whether this holds true for Mandarin speakers as well. That is, the experiments use stimuli chosen from all three languages in tests of native listeners in order to see how listeners of different languages carve up the perceptual space.

A second objective involves differences across age groups. As shown in both Chapter 4 and Chapter 6, children's early productions are mostly intermediate and less clear. Thus, it is hypothesized that adult will be less consistent in categorizing children's productions than in judging adults'. When evaluating children's early productions, previous studies have emphasized the variability that children's productions show, but few studies seem to address the question of how variable adults' judgments could be in listening to children' speech. If adults are less cosistent in judging children's productions, the reliability of the transcription method should be questioned.

A third objective relates to possible confound in the experiment with Japanese-speaking listeners reported in Chapter 5. Recall that the Japanese perception experiments reported in Chapter 5 presented an unexpected result. Specifically, the 20 Japanese listeners disagreed with each other when judging adult fricative stimuli. In Section 5.3.2, it was speculated that one possible explanation for this result is that the adult stimuli were produced by speakers from different dialect regions in Japan, and the inconsistency that was seen in Figure 5.2 on page 81 may be due to the differences in production patterns for the two fricatives in different dialects. The current experiments aim to eliminate this confounding factor by selecting adult stimuli from speakers of the Tokyo dialect.

7.2 Methods

7.2.1 Stimuli

The stimuli are extracted consonant-vowel sequences from productions of children's word repetitions that were described in Chapter 6 and adults' productions described in Chapter 3, which constitute a database of three languages that contains productions from children aged 2 to 5, with 10 children for each age group, as well as 10 adults,

with gender being balanced for each age group. Before stimulus selection, children's segmented word files in the database were carefully screened so that those of poor recording quality owing to background noise or otherwise distorted or masked sound, or those whose vowels are too short (such as /sɪstɚ/ 'sister' in English and /sun.wu.kõ/ 'monkey king' in Mandarin) were excluded during the process of stimuli selection.

Since chapter 4 has shown that for both English-speaking and Japanese-speaking listeners, M1 and onset F2 are the two most relevant acoustic dimensions, the stimuli were then selected to resemble the original distributions of M1 and Onset F2 in the fricative productions in the database. More specifically, the stimuli were selected to mirror the probability distributions of M1 and onset F2 respectively. Figure 7.1 compares the distributions of M1 and onset F2 of all the productions in the database (in gray bars) with the distributions of these two parameters in the selected stimuli (in striped bars) for all three languages, and it is clear that the selected stimuli have a range of M1 and onset F2 values that are representative of the distributions of these parameters in the database. Further, speaker age is controlled by having an equal number of stimuli for each age group with gender being balanced in both children and in adults. For each age group, both accurate productions and productions of language relevant error types were included, with the younger age group having more error tokens included than older age groups to mirror the error distribution in different age groups of the database. Moreover, the vocalic context effect is controlled by selecting equal number of stimuli for each consonant-vowel combination for both children and adults. There are a total number of 250 English stimuli, 240 Japanese stimuli and 500 Mandarin stimuli. These stimuli were blocked by language, and thus constituted three blocks in total. The reason to block by language instead of mixing them together is to prevent the subjects from applying nonnative perceptual criteria in judging their native language.
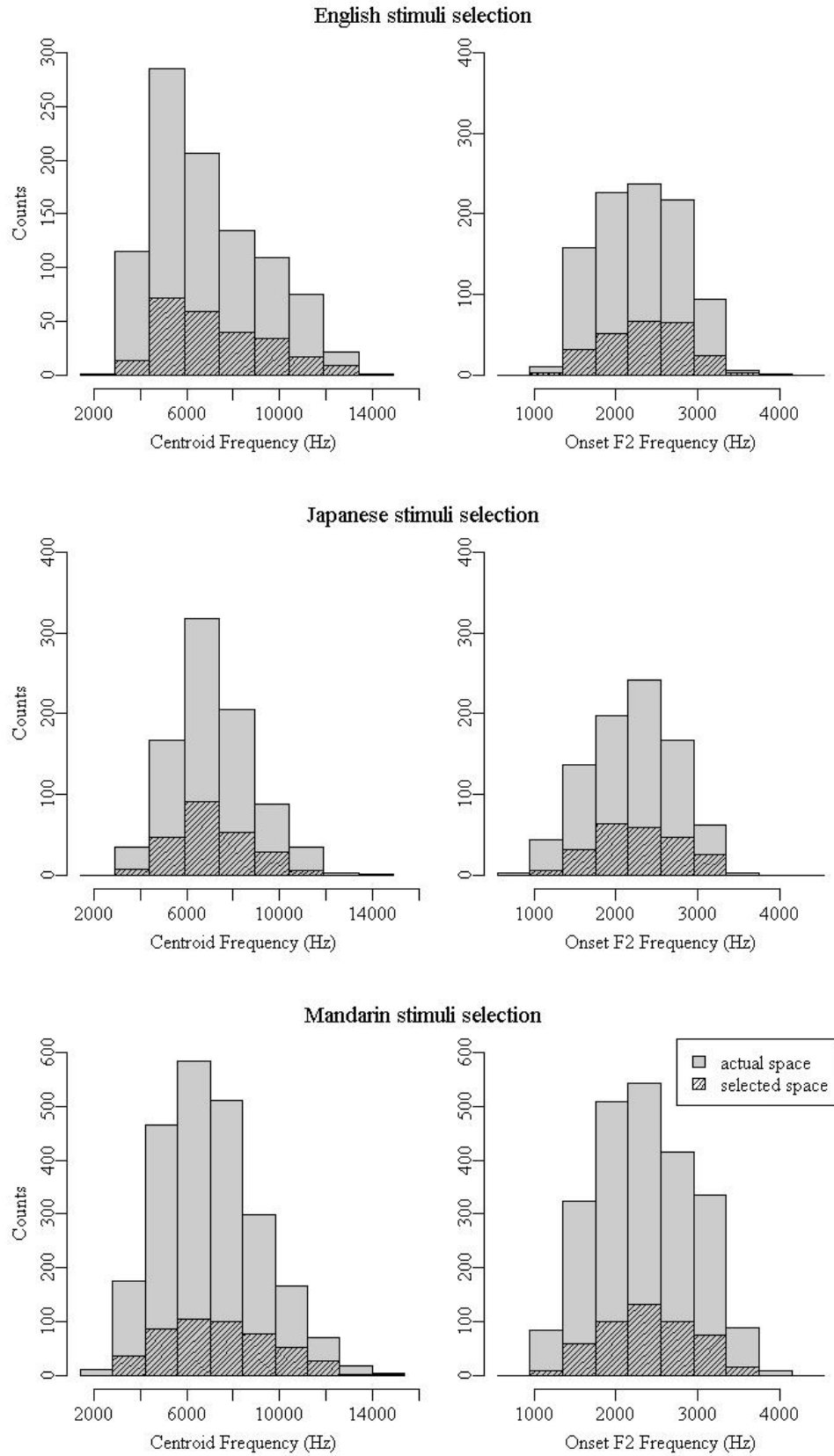
Figure 7.1: Comparison of the original distribution of M1 and onset F2 in the database with the selected distribution for the three languages

7.2.2 Participants and Task

Sixty participants aged 18 to 30 were tested in their home countries. That is, 20 English-speaking subjects were tested in Minneapolis, MN, 20 Japanese-speaking subjects were tested in Tokyo, Japan, and 20 Mandarin-speaking subjects were tested in Songyuan, China. All participants attended three sessions, each containing one block of stimuli for each of the three languages. That is, they heard one block of English, one block of Japanese, and one block of Mandarin in three different sessions. All the stimuli in each block were randomized for each subject. The presentation orders of the blocks were counterbalanced in the participants.

Participants were engaged in a task that required them to make a two-step judgment for each of the CV sequences they heard. For the first step of a trial, upon hearing a CV stimulus, they were forced to make a category choice by pressing a button on the computer keyboard. The English- and Japanese-speaking listeners chose between the s and h keys. The labels s and h correspond to the fricative categories /s/ ('s') and /ʃ/ ('sh') in English. Mandarin-speaking listeners were instructed to choose among 's' (/s/), 'sh' (/ʂ/), and 'x' (/ɕ/) by pressing the s, h or x respectively. For the second step, the same CV stimulus was presented again, and the participants were asked to rate the goodness of the CV initial fricative in relation to the category already selected, using the direct magnitude estimation method. That is, they were instructed to form a base number on their own for the goodness rating of the first stimulus, and then compare the goodness of the following stimulus with that of the first one. If the second stimulus is twice as good as the first one, the number they assigned for the second stimulus should be twice as big as the base number.

During the experiment, each participant, wearing a set of Sennheiser Closed Circumaural Headphone (HD 280 PRO), was seated in front of an IBM Thinkpad

laptop, and was given a numeric keypad to facilitate number entering for the goodness rating portion. Before each session started, participants were asked to listen to a few practice items to get familiar with the task, and they were given time for a break after every 100 trials.

Since this set of experiments is part of a larger project that has bigger goals, only part of the data are included and analyzed to answer the specific questions related to this dissertation. Specifically, this chapter only analyzes the forced-choice data, but not the direct magnitude estimate data on goodness rating. Also only the data on listeners' judgments on their native language are included, since the purpose is to find out how adult listeners react to the productions by children and adults in their own language.

7.3   Data analysis and results

7.3.1   Language-specific perceptual norms for the three languages

The procedure for analyzing language-specific categorical perceptions is similar to the one introduced in Chapter 5, where the community opinion of whether a particular stimulus item is s-like or sh-like is calculated by dividing the total number of 's' judgments (or 'sh' judgments) by the total number of subjects. And then, (using the binomial function to calculate how many out 20 listeners judgments need to agree to be above chance) tokens with more than 14 listeners's positive responses are considered as consensus 's' or consensus 'sh' (or consensus 'x' for Mandarin listeners) from the community's point of view.

Logistic regression models were fit for each of the language to answer two questions: 1.What are the language-specific perceptual norms in terms of the acoustic dimensions that are crucial in categorizing fricative contrasts? 2.Will listeners perceive adults' and childrens' productions differently?

More specifically, for languages that have a two-way contrast (i.e. English and Japanese), the dependent variable for the model is the two perceived fricative categories, that is, 'consensus 's'' (coded as '0') and 'consensus 'sh'' (coded as '1') as judged by the 20-listener community. The independent variables are as follows: normalized acoustic values for M1, M2, and onset F2, a categorical variable 'age' which has two levels :children vs. adults, as well as the interaction between 'age' and the three acoustic parameters.

Perceptual correlate	Coefficient	SE	Wald Z	p
Intercept	-7.4	5.2	-1.4	0.154
M1	-12.3	6.2	-2.0	**0.048**
M2	-1.5	1.5	-1.0	0.316
Onset F2	0.4	1.5	0.3	0.795
Age group (children vs. adults)	6.3	5.2	1.2	0.226
Age group & M1	6.3	6.3	1.0	0.317
Age group & M2	1.5	1.5	1.0	0.310
Age group & Onset F2	1.1	1.6	0.7	0.500

Table 7.1: Summary of the logistic regression model for the English-speaking listeners perception of /s/-/ʃ/ contrast.

Table 7.1 lists the results of the regression model for English, where the dependent variables are /s/ (the 's' category, coded as '0') vs. /ʃ/ (the 'sh' category, coded as '1'). It is clear from the table that only M1 is correlated significantly with the community's categorical judgment of the /s-ʃ/ contrast. Neither M2 nor onset F2 are significant in this model. Also in terms of the explanatory power, M1 has an overriding contribution than the other two parameters, as shown by the much bigger

absolute value of the coefficient for M1 (*-12.3*) than that for M2 (*-1.5*) and that for onset F2 (*0.4*). Another thing to note is that there is no overall age effect in the model, indicating that English listeners are not biased in one or the other fricative category for children. There is no significant interaction of age with any of the three acoustic parameters, suggesting that listeners do not use any of these three perceptual correlates differently in judging children's from adults' productions.

Perceptual correlate	Coefficient	SE	Wald Z	p
Intercept	-1.5	0.3	-3.0	0.154
M1	-1.8	0.5	-3.9	**<0.001**
M2	-1.3	0.4	-3.3	**<0.001**
onset F2	1.0	0.4	2.7	**<0.001**
Age group (children vs. adults)	-2.9	2.5	-1.2	0.245
Age group & M1	-5.1	3.2	-1.6	0.110
Age group & M2	1.4	1.5	0.9	0.370
Age group & Onset F2	0.2	1.2	0.1	0.884

Table 7.2: Summary of the logistic regression model for the Japanese listener's perception on /s-ɕ/ contrast.

Table 7.2 reports the results of the logistic regression model on Japanese-speaking listeners' perception. Different from the perceptual pattern for English-listeners, where only M1 correlates significantly with fricative categorization, all three acoustic parameters, M1, M2 and onset F2, correlate with Japanese listener's fricative perception. M1 contributes the most to the model, but the difference between the size of the coefficients is much smaller. M2 ranks second in terms of explanatory power, and onset F2 is third. In addition, no significant effects were found for age or for the interactions between age and any of the three acoustic parameters.

Two logistic regression models were fit in the Mandarin perception data. One was to find out the perceptual correlates that are relevant for the place contrast between /s/ and /ʂ/ in Mandarin, and the other is for the palatalization contrast

Perceptual correlate	Coefficient	SE	Wald Z	p
Intercept	9.0045	2.0802	4.3	**<0.001**
M1	-0.0001	0.0001	-1.5	0.121
M2	-0.0007	0.0005	-1.4	0.162
onset F2	-0.0034	0.0007	-4.8	**<0.001**
Age group (children vs. adults)	-3.7667	2.3100	-1.6	0.103
Age group & M1	0.0002	0.0001	1.9	0.055
Age group & M2	0.0009	0.0005	1.6	0.101
Age group & Onset F2	0.0006	0.0007	0.8	0.414

Table 7.3: Summary of the logistic regression model for the Mandarin listener's perception on palatalization contrast between /s,ʂ/ and /ɕ/

between /s, ʂ/ and /ɕ/. For the first regression model, the independent variables are /s/ (the 's' label) and /ʂ/ (the 'sh' label), and the dependent variables are the same with the Japanese and the English models, which include M1, M2, onset F2, age as well as interactions between age and the three acoustic parameters. The results show that M1 itself is able to explain all the variability in the model, and no other independent variables are needed. For the second regression which tests the perceptual norms of the palatalization contrast, /s/ and /ʂ/ were grouped together to form one 'nonpalatalized' variable, which is coded as '1', and /ɕ/ itself forms another 'palatalized' variable, which is coded as '0'. The dependent variables are then 'palatalized' vs. 'nonpalatalized' categories. The independent variables are the same as in the other models introduced in this section. Table 7.3 lists the results of the palatalization model. It is clear from the table that only onset F2 is significantly correlated with the percept of palatalization that differentiates /ɕ/ from the other two fricatives. Again, no significant age effect was found.

To summarize, clear crosslinguistic perceptual differences were found in the current experiments, where English listeners use only M1 in their fricative categorization, whereas Japanese listeners use M1, M2 and onset F2. Mandarin listeners use M1 in perceiving the place contrast, but onset F2 in the palatalization contrast. These perception results parallel remarkably with the production results on adults' productions discussed in Section 3.5 in Chapter 3. Especially for English and Japanese, the production and perception patterns are exactly the same in terms of the acoustic parameters that were found to be significant. For the Mandarin speakers, the production and perception patterns coincide exactly for the place contrast (only M1 being significant) and diverge for the palatalization contrast only in that M1 did not contribute significantly to the perception of /ɕ/ versus /s/ and /ʂ/.

Comparing the results on the English- and Japanese-speaking listeners in this chapter with those for the perception experiments in Chapter 5, there are a few differences as well as some similarities. Specifically, onset F2 was shown to be significant together with M1 in predicting fricative identities for English listeners in Chapter 5, whereas only M1 is significant in the current experiments. Also, M2 was not significantly predictive of fricative categories for the Japanese-speaking listeners in Chapter 5, whereas it is not only significantly correlated with fricative perception in the current experiment, but also contributes more than onset F2 in the model. The different statistical results may be owing to the different tasks, where the previous experiments ask listeners to respond 'yes' or 'no' to the questions of 'Is is an 's'?' and 'Is this an 'sh'?' in different blocks, the current experiment is a forced-choice between the two fricatives in a single block. Also where previous experiments included stimuli from both languages, in these experiments, the stimuli are blocked by language and only the responses on the native language stimuli are reported.

Despite these differences, the two sets of experiments also share some similar patterns. For example, English listeners consistently use M1 as the primary perceptual correlate in their fricative perception, which mainly reflects the frequency range that the bulk of the energy falls into in the fricative spectrum. By contrast, Japanese listeners use other spectral cues, such as M2 or onset F2 similarly as M1. So the results of both studies suggest that Japanese listeners attend to cues differently in differentiating contrastive fricatives than English listeners.
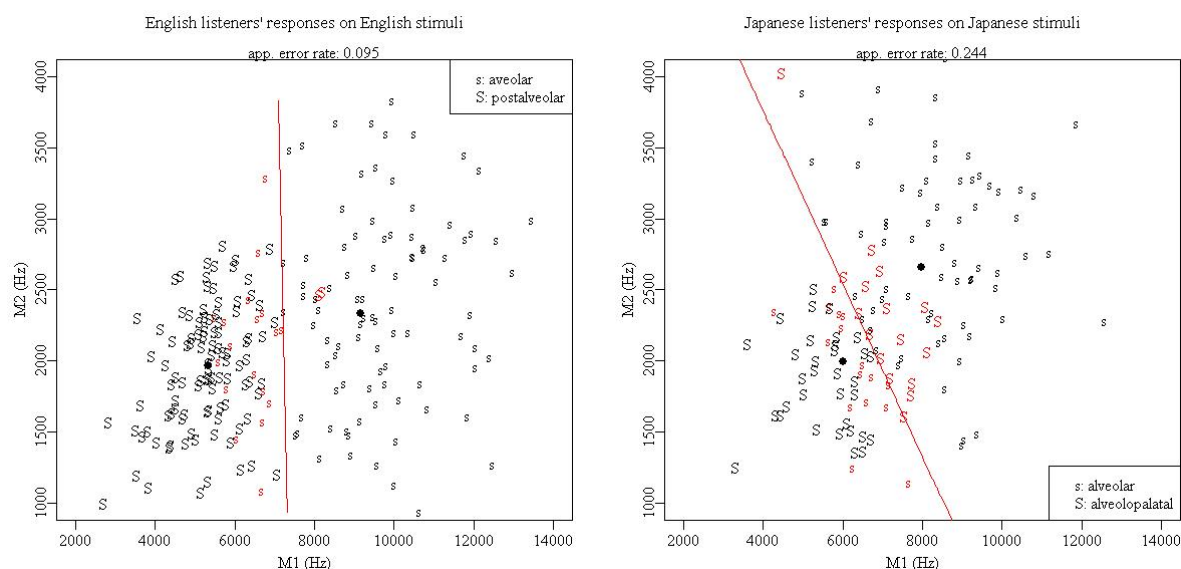
Figure 7.2: Comparison of perceptual norms between English and Japanese in the M1 by M2 space.

Figure 7.2 compares the English-speaking listeners responses with those of the Japanese-speaking listeners in the M1 and the M2 dimensions. For both graphs, linear discriminant function curves are drawn. It is clear that for English listeners, M1 is the primary perceptual cue, whereas M2 contributes little to the discrimination of the two categories, as shown by the almost vertical discrimination function line.

114

The pattern is different for Japanese, however, where M1 and M2 jointly discriminate the two categories.

Figure 7.3 plots listeners' perceptual division of the contrastive fricative categories in the M1 by onset F2 space for the three languages. In these set of graphs, symbol 's' representing those stimuli that were judged by at least 14 listeners to be /s/ in all three languages. Symbol 'S' is for those were judged by at least 14 listeners to be /ʃ/ by English listeners, consensus /ɕ/ by Japanese listeners, and consensus /ʂ/ by Mandarin listeners. Symbol 'x' is for those correctly judged as /ɕ/ by Mandarin listeners only. It is clear from comparing the three graphs that listeners of the three languages carve up the perceptual space differently. With English listeners, M1 is mainly used in the discrimination of the /s-ʃ/ contrast, with onset F2 contributing very little. By contrast, Japanese listeners divide more on onset F2 in their discrimination of the /s-ɕ/ contrast. Further, Mandarin listeners dissect the same perceptual space into three, where the stimuli having lower onset F2 values were judged as either /s/ or /ʂ/ depending on their M1 values, with higher M1 for /s/ and lower M1 for /ʂ/, and the higher values of onset F2 were judged as /ɕ/. These perceptual patterns replicate what has been found in Chapter 5.

7.3.2   Inter-listener agreement

7.3.2.1   Percentage of consensus identification on target fricatives produced by adults

One issue that emerged from the analysis of the perception data in Chapter 5 is that Japanese-speaking listeners behave differently from English-speaking listeners in terms of accuracy and the degree of agreement in judging adult stimuli. In particular, English listeners were very accurate in identifying the target fricative in adult stimuli,
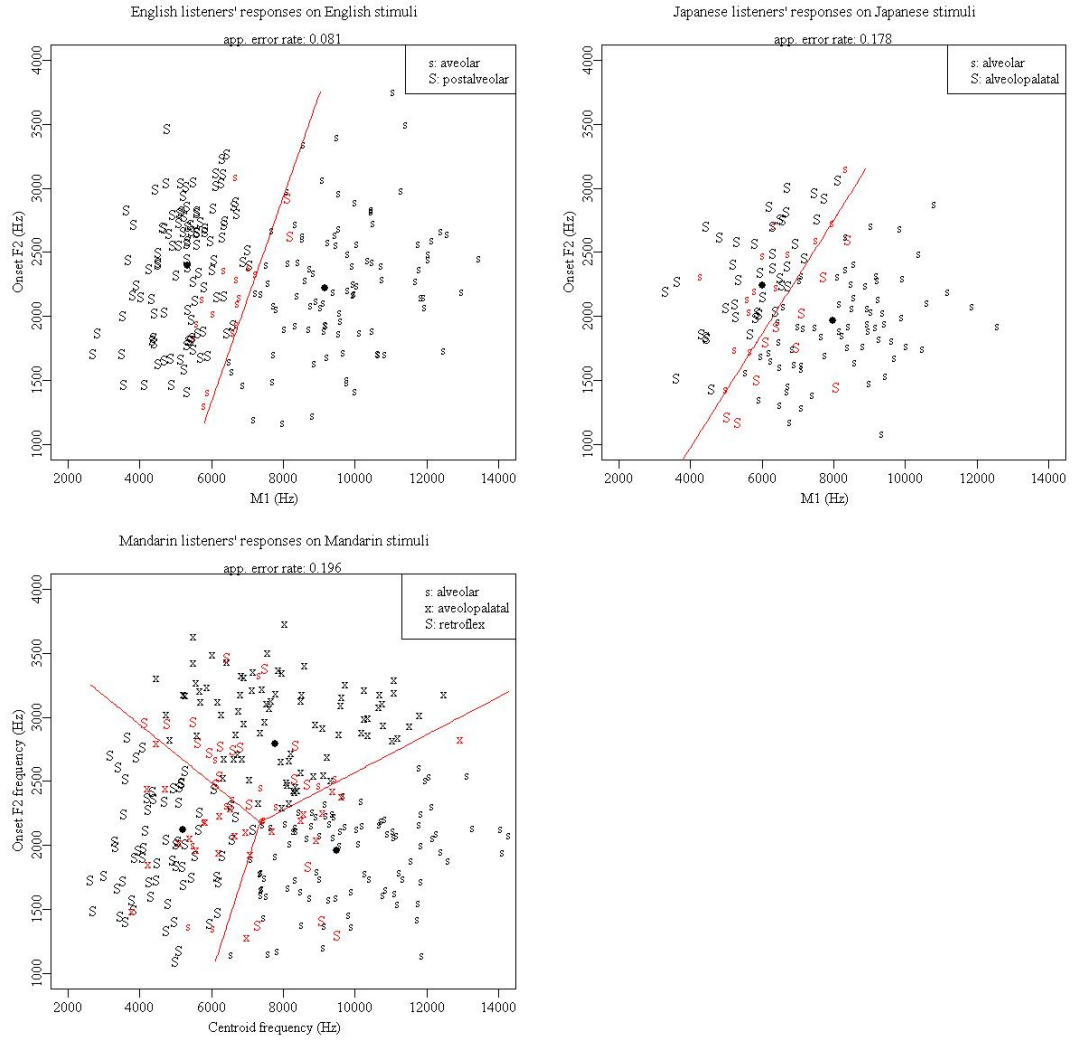
Figure 7.3: Comparison of perceptual norms across the three languages in the M1 by onset F2 space.

and the 20 listeners agreed with each other for the most part. Japanese listeners, however, not only did not achieve the same degree of accuracy as English listeners in identifying adult stimuli, but also disagreed with each other more in the identification. This is unexpected since adults' speech was considered as clear and unambiguous, and should not have posed any problem for the listeners to identify. In Chapter 5, several possible explanations for these unexpected result were suggested, one of which is that all of the Japanese listeners were speakers of the Tokyo dialect, whereas some of the adult stimuli they were asked to identify in the experiments were produced by speakers of other dialects. The low rate of accuracy and agreement may come from dialect differences in production of sibilant fricatives. This possibility is explored further in the current section, since the problem of dialect effects have been corrected by including adult stimuli from only the Tokyo dialect in the current experiments.

Figure 7.4 plots the percentage of correct identification for listeners of the three languages when they listen to the adults' productions of their own language. It shows a similar pattern to that in Figure 5.2 on page 81 in that English-speaking listeners are very consistent in identifying the target sibilants, whereas Japanese listeners still agree much less than do English-speaking listeners for both targets. Since in the current experiments, all 240 stimuli were produced by Tokyo speakers, this difference in listener agreements in judging adults' productions cannot be attributed to the dialect differences. There were three other possibilities discussed in Chapter 5. One is that the confusion may come from the Japanese writing system. More specifically, the Japanese writing system mixes phonographic hiragana and katakana with logographic kanji (Chinese characters). The hiragana and katakana graphemes are a syllabary, in which each graph or digraph represents a moraic segment or segment sequence. Figure 7.5 lists the hiragana graphemes in the goju-on arrangement that Japanese speakers were taught in elementary school. (This arrangement also defines the order in
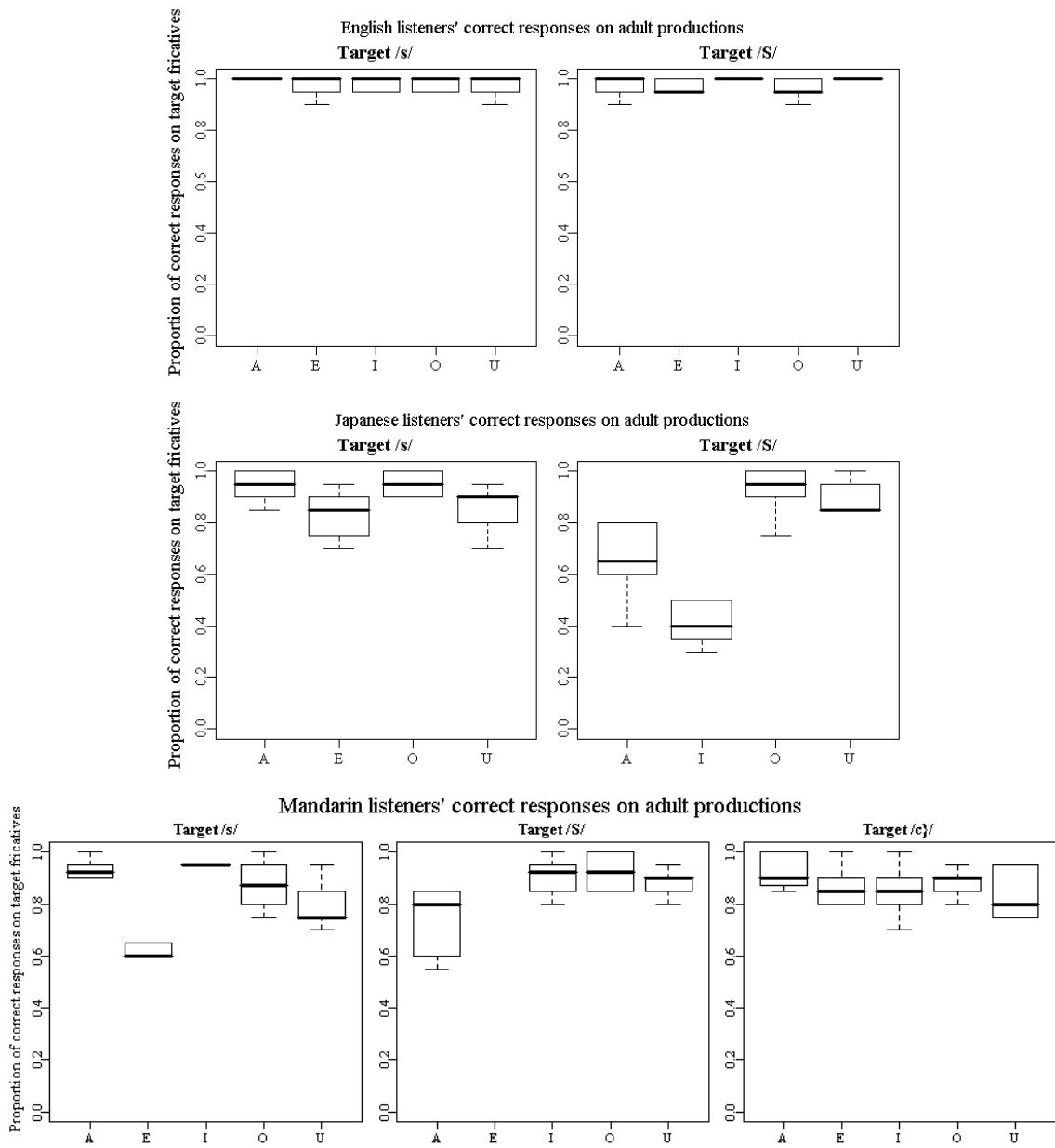
117

Figure 7.4: Comparison of adults' perception of targets on adults' productions for different vowel context across the three languages

which words are listed in dictionaries, telephone books, and so on.) The bottom part of Figure 7.5 is the "appendix" that teaches first graders how to read and write the digraphs for sequences such as /ɕa/, /ɕu/ and /ɕu/. (In a dictionary, words beginning with these sequences come immediately with words beginning with /ɕija/, /ɕijo/ and /ɕiju/.) The Japanese writing system thus teaches a metalinguistic awareness of the diachronic source of phonotactic distribution, whereby /ɕi/ was historically an allophone of /s/ before /i/, and alternates with it in the inflectional morphology. Children are taught /ɕi/ as part of the /s/ row of the goju-on arrangement.

However, even if the goju-on arrangement can explain why the accuracy and agreement is particularly low for the /ɕi/ sequence, the other sequences should not be affected, and Figure 7.4 suggests that the other sequences in Japanese still do not achieve the same kind of accuracy and agreement levels as do /s/ and /ʃ/ for the English-speaking listeners. Thus, another possibility is that because of the opaque relationship between the Japanese writing system and the sounds, Japanese listeners are not trained to recognize phonemes as much as English-speaking listeners are. This possibility can actually be tested in Mandarin. Even if Mandarin has a logographic writing system, people are trained to learn the romanization alphabet (Pinyin) in elementary school as a way to learn the sounds of the Chinese characters. Therefore they are well drilled to 'spell' the sounds of morphemes written by given Chinese characters and are tested on their exams throughout school education till high school. For example, they were taught that Pinyin letter 's' is for the sound /s/, the digraph 'sh' is for /ʂ/, and 'x' for 'ɕ'. So even if the three fricatives do not share complete overlapping distributions as the Japanese pair, Mandarin listeners were trained to categorize them separately through Pinyin. When looking at the Mandarin listeners' responses in the bottom panel of Figure 7.4, it is clear that Mandarin listeners do not achieve the same level of accuracy and agreement as the English subjects. Therefore,

C	a	i	u	e	o
–	あ	い	う	え	お
k	か	き	く	け	こ
s	さ	し	す	せ	そ
t	た	ち	つ	て	と
n	な	に	ぬ	ね	の
h	は	ひ	ふ	へ	ほ
m	ま	み	む	め	も
j	や		ゆ		よ
r	ら	り	る	れ	ろ
w	わ				を
	ん				

きゃ		きゅ		きょ
しゃ		しゅ		しょ
ちゃ		ちゅ		ちょ
にゃ		にゅ		にょ
ひゃ		ひゅ		ひょ
みゃ		みゅ		みょ
りゃ		りゅ		りょ

Figure 7.5: Goju-on arrangement in hiragana (courtesy of Dr. Mary Beckman). Sequences with /ɕ/ initials are highlighted in gray.

the possibility that the low accuracy in the other sequences of Japanese cannot be attributed to the opaqueness of the writing system either.

The third possibility discussed in Chapter 5 is the frequency effects in determining the accuracy rating in adult listeners. Although this explains the low rating for /se/(extremely low-occuring sequence in the Songyuan Mandarin) in Mandarin and /ɕa/ in Japanese, it does not get support from the rest of the sequences. Moreover, it cannot explain the crosslinguistic differences in general found between English and Japanese/Mandarin.

This only leaves the fourth suggested explanation that the phonological representations of the fricative categories in English are more robust or clearcut than

those in Japanese and Mandarin. This may be because of the phonological distributions. The English fricatives contrast in front of all vowels, whereas the Japanese and Mandarin fricative contrasts are neutralized in some vowel contexts. It may be also because of differences in the writing system where the alphabet of English reinforce the categorizations in English, but not so in Japanese and Mandarin. Even if Mandarin speakers were trained to use the phonetic alphabet, it is not part of their native writing system, and is only of use during their learning of characters. In order to tease apart these two possible sources, either English and Japanese listener's fricative perception must be tested using a different task that does not require their metalinguistic phonological knowledge or their perception must be tested on a different set of sounds where the English pair has the partial overlapping distribution, whereas the Japanese pair share complete overlapping distribution. This needs to wait for future research.

7.3.2.2   Listener agreement as a function of speaker age

In this section, I test a prediction about perception that is suggested by the patterns shown in Figure 6.3 on page 98. Specifically, I predict that adults will be less certain and less consistent in categorizing children's speech than adults' speech and that they will be specifically inconsistent for the youngest children.

Figure 7.4 showed one way to evaluate agreement between listeners by calculating how many of the 20 listeners agree with the target. This works for the purpose of examining listeners' agreement on adults' productions, since the target category can be assumed to be the same as the perceived fricative category in most cases. When evaluating listeners' agreement regarding children's productions, however, this measure does not work, since adults may agree with each other about which fricative category they perceived a child to have produced, but the perceived category may not

be the target category. For example, adults frequently perceive children to substitute one sound for another, which does not mean that they do not agree with each other, but rather that they agree with each other in that the perceived category is different from the target one. Since the question of interest here is to see whether listeners agree with each other regardless of whether the agreed category is the same as the target one, another way of assessing agreement has to be employed.

A common way of evaluating inter-transcriber agreement/reliability between two judges on children's productions without referencing the target sound is Cohen's kappa (Byrd, Conture, and Ohde, 2007; Neuman, Koh, and Dwyer, 2008). Cohen's kappa (Cohen, 1960) is a statistic that measures how much two raters agree with each other when each of them makes mutually exclusive categorical judgments on a fixed number of items. It is a conservative measure since it takes out the amount of agreement introduced by chance. The value of Cohen's kappa ranges between -1 to 1, where 1 indicates perfect agreement between the two judges, and $<= 0$ means no agreement. The current experiments are exactly the type of task that Cohen's kappa could evaluate: they involve listeners' judgments to classify any given CV stimulus into two mutually exclusive categories – 's' or 'sh', or in the case of Mandarin into three categories – 's', 'sh' or 'x', on a total of several hundred of stimuli. The only problem with Cohen's kappa is that it only deals with measuring agreement between two raters while each of the current experiments has 20 listeners/raters. Fleiss's kappa (Fleiss, 1971, 1981) is then selected to use since it is exactly the same as Cohen's kappa, but can measure agreement rates among multiple raters.

More specifically, for the current experiments, Fleiss's kappa was calculated over 20 listeners on their categorical responses of whether a given stimulus is a consensus 's' or 'sh' in English and Japanese. For Mandarin, it was calculated over 20 listeners' responses on three categories – 's', 'sh' or 'x'. Similar to Cohen's kappa,

Fleiss' kappa also has a value between -1 and 1 with 1 for perfect agreement and <= 0 for disagreement.

Table 7.4 presents the values of Fleiss' kappa statistic on 20 English listener's responses on English stimuli for each age groups and for each vowel context. It is clear from the table that the inter-listener agreement increases as the stimuli-producers' age increases. When listening to the 2-year olds' productions, the 20 listeners achieve an agreement less than 0.5 (0.49), as compared with the agreement rate for adults' productions, 0.88. When the Fleiss' kappa is decomposed into different vocalic context, one thing that is immediately noticeable is that the inter-listener agreement for the rounded vowels (/o/ and /u/) is lower than that for the unrounded vowels for all age groups except the adult group. This suggests that listeners tend to disagree with each other when children produce sibilant fricatives in front of rounded vowels. This may be owing to the fact that by protruding the lips, the length of the front cavity is comparatively lengthened, and therefore scales down the overall frequency range that the major energy concentrates. For adult speakers' productions, though, the /s-ʃ/ differences are made so clear and so robust that even the rounding would not introduce confusion. For children's productions, however, the two contrastive categories were not produced as clearly separate as in the adult's cases, plus there are more variabilities in children's productions on both fricative categories. Therefore, the rounding effect in children's productions may pose more problems for listeners to normalize than in adults' productions. Alternatively, this may reflect the lower frequencies for these contexts in English, as shown in Table 2.2 in Chapter 2, where both /s/ and / ʃ/ have lower frequencies in the context of /o/ and /u/ than in other vowel contexts.

Table 7.5 presents the inter-listener agreement for the 20 Japanese listeners' responses on the Japanese stimuli. Compared to table 7.4, the inter-listener agreement

Stimuli age group	Vocalic context				
	A	E	I	O	U
2;0 - 2;11	0.64	0.44	0.58	0.36	0.42
3;0 - 3;11	0.73	0.73	0.78	0.56	0.38
4;0 - 4;11	0.62	0.57	0.49	0.53	0.44
5;0 - 5;11	0.85	0.88	0.86	0.63	0.68
adults (18 -30)	0.92	0.84	0.86	0.85	0.92

Table 7.4: Fleiss's kappa for 20 English-speaking listeners judgments of 's' or 'sh' in stimuli produced by English speakers of different age group and in different vocalic contexts.

score is strikingly low. The agreement on adult's productions is already generally low, which reflects the low degree of consensus shown in Figure 7.4 discussed in the previous section. Further, listeners' agreement on children's productions is even lower, which is expected since children's productions are less clear and less mature, and the pattern of the agreement rating being increased with age is not as clear as in the English experiments. Moreover, the agreement ratings are especially low in the context of vowel /e/ and /i/ for both children and adults. This is expected as well since these are contexts where the two sibilant fricatives in Japanese do not contrast or only contrast marginally. (As mentioned before, /si/ is phonotactically illegal in Japanese, and /ɕe/ only marginally occurs.)

Table 7.6 lists the Fleiss's kappa statistic for the Mandarin-speaking listeners' judgments of stimuli produced by children of each age groups as well as by adults. Generally speaking, listeners' agreement is very low for age 2-3; most of the time, the kappa value is below 0.25. The agreement becomes much higher in the age groups 4 and 5, where half of the 20 listeners agree with each other most of the time, except in the vowel /e/ context. The vowel /e/ is the context where /s/ and /ʂ/ rarely occur,

Age group	Vocalic context				
	A	E	I	O	U
2;0 - 2;11	0.48	-0.05	-0.002	0.39	0.20
3;0 - 3;11	0.56	0.270	0.28	0.41	0.19
4;0 - 4;11	0.24	0.16	0.04	0.60	0.41
5;0 - 5;11	0.48	0.10	0.03	0.57	0.31
Adults	0.40	0.01	0.08	0.77	0.57

Table 7.5: Fleiss's kappa for 20 Japanese-speaking listeners judgments of 's' or 'sh' in stimuli produced by Japanese speakers of different age group and in different vocalic contexts.

Age group	Vocalic context				
	A	E	I	O	U
2;0 - 2;11	0.17	0.22	0.25	0.14	0.17
3;0 - 3;11	0.06	0.18	0.27	0.25	0.20
4;0 - 4;11	0.52	0.19	0.44	0.38	0.44
5;0 - 5;11	0.48	0.17	0.52	0.42	0.56
Adults	0.68	0.22	0.68	0.66	0.55

Table 7.6: Fleiss's kappa for 20 Mandarin-speaking listeners judgments of 's' or 'sh' in stimuli produced by Mandarin speakers of different age group and in different vocalic contexts.

it is not just on children's productions; listeners' agreement is low for adults' stimuli as well.

7.4   Summary and conclusion

In sum, listeners of different languages do not carve up the perceptual space in the same way. English listeners rely primarily on M1, whereas Japanese listeners rely on more acoustic cues including M1, M2 and onset F2. Mandarin listeners attend to M1 in perceiving fricatives with different places of articulation, but to onset F2

in categorizing fricatives that differ from each other in tongue posture (palatalized posture vs. nonpalatalized posture). Furthermore, the adult perceptual patterns of voiceless sibilant fricatives in the three languages have been shown to parallel the adult production patterns in Chapter 3, as well as children's acquisition patterns in Chapter 6.

In addition, whether the speaker who produced the stimulus was an adult or child did not significantly contribute to the logistic regression, although the coefficient for this independent variable was very large for the English-speaking listeners that suggested many more /s/ judgments for child productions. There does exist another perceptual difference, however, that relates to speaker's age. That is, listeners are far less likely to reach a consensus when judging children's productions as compared with adults', as gauged by Fleiss's Kappa. Moreover, the younger the speakers' age, the harder it was for multiple listeners to get high agreement on the stimuli. In addition, listener agreement was also affected by phonological distributions. More specifically, in the vocalic contexts where fricatives do not contrast, they have more disagreement.

CHAPTER 8

CONCLUSION AND DISCUSSION

This dissertation is an attempt to look for the phonetic universals in children's frica-
tive development, since the search for phonological universals has failed or been con-
tradictory in crosslinguistic comparisons in previous studies. The main reason it fails,
in my point of view, is that phonological description of children's productions, which
mainly employs the transcription method, is not a reliable depiction of children's
early productions as it requires adults to make categorical judgments using language-
specific criteria. In order to avoid the distortion of using the transcription method,
acoustic analysis is used as the main tool in this study to examine children's produc-
tions and compare them across languages. Also a series of perception experiments
were performed in order to evaluate whether adults' perceptual norms are different
across languages, and if so, in which aspects.

There are four major findings in this dissertation. First, acoustic analysis on
adults' productions discussed in Chapter 3 revealed that adult speakers of different
languages do not share the acoustic dimensions as far as how fricative categories
are differentiated. While English speakers mainly produce /s/ and /ʃ/ differently in
the M1 dimension, Japanese speakers produce /s/ and /ɕ/ consistently different in
more acoustic dimensions, including M1, M2 and onset F2. Mandarin speakers make
their place distinction between /s/ and /ʂ/ in the M1 dimension, and palatalization
distinction between /s,ʂ/ and /ɕ/ mainly in the onset F2 dimension.

Second, both Chapter 5 and Chapter 7 showed that adult listeners of different languages also attend to different acoustic cues in identifying adults' and childrens' fricative productions. Moreover, the perceptual norms parallel the production norms for adult speakers of these three languages. That is, not only English listeners differentiate their /s/ and /ʃ/ productions in the M1 dimension, but also their perceptions primarily correlate with M1 in categorizing fricatives produced by children as well as by adults. Similarly, Japanese listeners attend to more acoustic cues, mainly including M1, M2 and onset F2, which are also the dimensions that speakers use most in making /s-ɕ/ distinctions in Japanese. These results suggest that adults both produce and perceive sibilant fricatives in a language-specific manner.

Third, it is found in Chapter 4 that children may be able to produce the contrastive fricative categories consistently differently, but just not in conformity with adults' perceptual norms. In other words, some children were found to produce fricatives contrastively in acoustic dimensions that adults do not use or attend to, and therefore were not able to be recognized as being able to produce the contrast 'correctly'. Since the adults' production and perceptual norms are different in different languages, the criterion of whether a particular contrast produced by children is 'correct' or not also depends on specific language. These results argue against using transcription as a way to describe children's productions crosslinguistically and especially to compare them across languages since transcription results described children's productions as filtered out by adult's language-specific perceptual criteria.

Last but not least, Chapter 6 uses acoustic analysis to describe children's productions. It was found that children start from an undifferentiated lingual gesture as evidenced by unimodal distribution for contrastive categories (in Figure 6.2 on page 95), no matter what language they speak. This unimodal distribution diverges into two categories in English and Japanese and into three categories in Mandarin.

More importantly, the relevant dimensions that the fricative categories contrast and the magnitude of the split are different across languages, and are consistent with the acoustic dimensions that adults produce and perceive these categories.

On the other hand, although children start from undifferentiated gestures in all three languages, such early forms are not identical across languages. In English, the initial fricative gesture generally has higher centroid frequency values (the M1 dimension), and it is not surprising that adult listeners would classify these productions into the /s/ category, even if they do not resemble the adult forms completely. In Japanese, the initial fricative productions generally have higher centroid frequencies and higher onset F2 values, which makes them easy to fall into the /ɕ/ category for Japanese listeners. The question that arises is why are children biased to produce different undifferentiated gestures in different languages? One possibility is that the input frequencies of these fricatives may play a role in shaping the initial stage of fricative acquisition. In English, /s/ is more frequent than /ʃ/ (at least in the adult lexicon), which indicates that English-speaking children are more likely to hear high-frequency fricative noise as in /s/ than low-frequency fricative as in /ʃ/. Similarly, in Mandarin, /s/ is least frequent among the word-initial voiceless sibilant fricatives in the adult lexicon, and it is the last one that was separated out from the unimodal distribution for Mandarin children. Another thing to note is that in English, /s/ also occurs in the word-final position as a bound morpheme to indicate plural form. It is possible that the high functional load of /s/ as a grammatical index can also reinforce the early production of /s/-like fricatives.

The frequency account does not predict the relative acquisition order in Japanese, where /s/ and /ɕ/ have similar frequency with /ɕ/ being slightly less frequent, but Japanese children were found to produce a /ɕ/-like form first. There are two potential explanations for this. One is that the frequency counts listed in Table 4.3 on page 58

129

are from an adult lexicon. There is anecdotal evidence suggesting that in Japanese child-directed speech, /ɕ/ is actually more frequent than /s/. The other possibility is related to the special vocal tract geometry in infants, whose oral cavities are much flatter than that of adult, since they do not have the erupted dentition (Bosma, 1975; Kent and Stark, 1981). When adults produce the alveolopalatal fricative, they have to form a narrow channel by raising the tongue body, but children do not need to - they have that narrow channel with the tongue in its resting position. It is likely that they can at least produce some kind of approximation closer to that of alveolopalatal target without much effort in the early stages.

Taken together, the acoustic and perception results in this dissertation suggest that there is a crosslinguistic general tendency in children's fricative development, for them to start from ambiguous non-differentiating category productions, and finely tune their category distinctions in accordance with adults norms as they gain more linguistic experience. This tendency may not be specific to fricative acquisition, and thus has larger implications as a more general cognitive process. That is, children are not born with categories as pre-specified, but rather form and shape their categories by attuning to how adults around them do. In addition to the language-universal tendency found, there are also some language-specific differences that were found in children's acquisition of fricatives. These differences might be accounted for by frequency effects; a possibility which awaits future investigations.

BIBLIOGRAPHY

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

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language.*, To appear.

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

Behrens, Susan and Sheila E. Blumstein. 1988. On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *Journal of the Acoustical Society of America*, 84(3). 861–867.

Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109(2). 775–794.

Bladon, Anthony, Christopher Clark, and Katrina Mickey. 1987. Production and perception of sibilant fricatives: Shona data. *Journal of the International Phonetic Association*, 17(1). 39–65.

Boersma, Paul and David Weenink. 2005. Praat: doing phonetics by computer.

Bosma, J. F. 1975. Anatomic and physiologic development of the speech apparatus.? D. B. Tower, editor, *The nervous system. Human communication and infants disorders.*, volume 3. New Haven, CT.: Raven.

Byrd, Courtney T., Edward G. Conture, and Ralph N. Ohde. 2007. Phonological priming in young children who stutter: holistic versus incremental processing. *American Journal of Speech-Language Pathology*, 16. 43–53.

CAO, YUN. 1986. Sex related phonetic variations in the beijing vernacular. *Chinese Studies.*, 6.

CATFORD, JOHN. C. 1988. *A practical introduction to phonetics.* Oxford: Clrendon Press.

CHAO, YUEN REN. 1948. *Mandarin primer.* Massachusetts: Harvard University Press.

CHAO, YUEN REN. 1968. *A Grammar of Spoken Chinese.* Berkeley: University of California Press.

CHO, TAEHONG AND PETER LADEFOGED. 1999. Variation and universals in vot: evidence from 18 languages. *Journal of Phonetics*, 27. 207–229.

COHEN, JACOB. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement.*, 20(1). 37–46.

DART, SARAH N. 1998. Comparing french and english coronal consonant articulation. *Journal of Phonetics*, 26(1). 71–94.

DARWIN, C. 1887. A biographical sketch of an infant. *Mind*, 2. 285–294.

DARWIN, CHARLES. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* London: John Murray, 1st edition. edition.

DELATTRE, P. C, A. M. LIBERMAN, AND F. S. COOPER. 1964. Formant transitions and loci as acoustic correlates of place of articulation in american fricatives. *Studia Linguistica*, 18. 104–121.

EDWARDS, J. AND M. E. BECKMAN. 2008. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in the acquisition of consonant phonemes. *Language Learning and Development*, 4(1). 122–156.

EINSTEIN, ALBERT. 1916. The foundation of the general theory of relativity. *Annalen der Physik*, 49.

EVERS, VINCENT, HENNING REETZ, AND ADITI LAHIRI. 1998. Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics*, 26(4). 345–370.

FERGUSON, C. A. 1986. Discovering sound units and constructing sound systems: It's child's play. J. S. Perkell and D. H. Klatt, editors, *Invariance and variability in speech processes*, 36–51. Hillsdale, NJ: Lawrence Erlbaum Associates.

FLEISS, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5). 378–382.

FLEISS, J. L. 1981. *Statistical Methods for Rates and Proportions.* Wiley series in probability and mathematical statistics. New York: John Wiley & Sons, 2nd edition.

FLETCHER, S AND D NEWMAN. 1991. [s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width. *Journal of the Acoustical Society of America*, 89. 850–858.

FORREST, K., G. WEISMER, M. HODGE, D.A. DINNSEN, AND M. ELBERT. 1990. Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clinical Linguistics and Phonetics*, (4). 327 – 340.

FORREST, K., G. WEISMER, P. MILENKOVIC, AND R. N DOUGALL. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, 84. 115–123.

FUNATSU, SEIYA. 1995. Cross language study of perception of dental fricatives in japanese and russian. *ICPhS 95*, 4. 124–127.

HALLE, MORRIS AND KENNETH N. STEVENS. 1997. The postalveloar fricatives of polish. Hajime Hirose Shigeru Kiritani and Hiroya Fujisaki, editors, *Speech Production and Language: In Honor of Osamu Fujimura*, volume 13 of *Speech 13*, 176–191. Berlin New York: Mouton de Gruyter.

HALLE, MORRIS, KENNETH N. STEVENS, BJORN LINDBLOM, AND SVEN E. G. OHMAN. 1979. Some reflections on the theoretical bases of phonetics. *Frontiers of Speech Communication Research*, 335–349. New York: Academic.

HAMANN, SILKE. 2003. *The Phonetics and Phonology of Retroflexes.* Ph.D. thesis.

HARRIS, KATHERINE SAFFORD. 1958. Cues for the discrimination of american english fricatives in spoken syllables. *Language and Speech*, 1. 1–7.

HEINZ, JOHN M. AND KENNETH N. STEVENS. 1961. On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 33. 589–596.

HEROARD, JEAN., editor. 1868. *Journal de l'Enfance et de la Premiere Jeunesse de Louis XVIII.* Paris: Firmin Didot Freres, fils & ci.

HIERONYMUS, JAMES L. 1994. Ascii phonetic symbols for the world's languages: Worldbet. *AT &T Bell Laboratories, Technical Memo*, 23.

HOCKETT, CHARLES F. 1960. The origin of speech. *Scientific American.*, 203. 89–97.

HU, MINGYANG. 1991. Feminine accent in the beijing vernacular: a sociolinguistic investigation. *Journal of the Chinese Language Teachers Association*, 26(1). 49–54.

HUANG, SHUDONG, XUEJUN BIAN, GRACE WU, AND CYNTHIA MCLEMORE. 1996. Callhome mandarin chinese lexicon.

HUGHES, G. W. AND M. HALLE. 1956. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, (28). 7.

INGRAM, DAVID. 1981. *Procedures for the phonological analysis of children's language.* Baltimore:MD: University Park Press.

IRWIN, CARTER A., J.R. AND M.J. BRIGGS-GOWAN. 2002. The social-emotional development of 'late-talking' toddlers. *Journal of the American Academy of Child and Adolescent Psychiatry.*, 41. 1324–1332.

JAEGER, T. FLORIAN. 2008. Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language.*, To appear.

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

JESUS, LUIS M. T. AND CHRISTINE H. SHADLE. 2002. A parametric study of the spectral characteristics of european portuguese fricatives. *Journal of Phonetics*, 30(3). 437–464.

JOHNSON, KEITH. 1997. *Acoustic and Auditory Phonetics.* Blackwell Publishers.

JONGMAN, ALLARD, RATREE WAYLAND, AND SERENA WONG. 2000. Acoustic characteristics of english fricatives. *Journal of the Acoustical Society of America*, 108(3). 1252–1263.

KEATING, PATRICIA A., CAROLE PARADIS, AND JEAN-FRANOIS PRUNET. 1991. Coronal places of articulation. *The Special Status of Coronals: Internal and External Evidence, II: Phonetics and Phonology*, 29–48. San Diego: Academic.

KENT, R. 1992. *The biology of phonological development.* Phonological development: Models, research, implications. Timonium, MD: York Press.

KENT, RAY 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.

KENT, RAYMOND D. 1976. Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 19.

KENT, RAYMOND D. AND RACHEL E. STARK. 1981. Articulatory-acoustic perspectives on speech development. *Language Behavior in Infancy and Early Childhood*, 105–126. New York: Elsevier North Holland.

KIER, W. AND K. SMITH. 1985. Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society*, 83. 307–324.

KONG, EUNJONG. 2009. *The Development of Phonation-type Contrasts in Plosives: Cross-linguistic Perspectives.* Ph.D. thesis, Ohio State University.

LADEFOGED, PETER. 1957. Use of palatography. *Journal of Speech and Hearing Disorders*, 22. 764–774.

LADEFOGED, PETER. 1989. Representing phonetic structure. *UCLA Working Papers in Phonetics*, 73.

LADEFOGED, PETER AND IAN MADDIESON. 1986. Some of the sounds of the world's languages. *UCLA Working Papers in Phonetics*, 64. 1–137.

LADEFOGED, PETER AND IAN MADDIESON. 1996. *The sounds of the world's languages.* Oxford, UK ; Cambridge, Mass.: Blackwell.

LADEFOGED, PETER AND ZONGJI WU. 1984. Places of articulation: An investigation of pekingese fricatives and affricatives. *Journal of Phonetics*, 12(3). 267–278.

LAMBACHER, S., W. MARTENS, B. NELSON, AND J. BERMAN. 2001. Identification of english voiceless fricatives by japanese listeners: the influence of vowel context on sensitivity and response bias. *Acoustical Science and Technology*, 22(5). 334–343.

LARIVIERE, CONRAD. 1975. The distribution of perceptual cues in english prevocalic fricatives. *Journal of Speech and Hearing Research*, 18.

LEE, WAI-SUM. 1999. An articulatory and acoustic analysis of the syllable-initial sibilants and approximants in beijing mandarin. *The International Congress on Phonetic Sciences*, 413–416. San Francisco.

LI, WEI, HUA ZHU, BARBARA DODD, FAY WINDSOR, M. LOUISE KELLY, AND NIGEL HEWLETT. 2002. Phonological saliency and phonological acquisition by putonghua speaking children: A cross-populational study. *Investigations in Clinical Phonetics and Linguistics*, 169–183. Mahwah, NJ: Erlbaum.

LINDBLAD, PER AND STURE LUNDQVIST. 1995. The groove production of swedish sibilants ?an epg analysis. *the 13th Internatinal Congress of Phonetic Sciences (ICPhS'95)*, volume 2, 458–461. Stockholm, Sweden.

LOCKE, JOHN L. 1983. *Phonological acquisition and change.* New York: Academic Press.

MACKEN, MARLYS A. AND DAVID BARTON. 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.

MAXWELL, E.M. AND G WEISMER. 1982. The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating childs voice contrast for stops. *Applied Psycholinguistics*, 3. 14.

MENN, L. AND BRIAN BUTTERWORTH. 1983. Development of articulatory, phonetic, and phonological capabilities. *Language Production, II: Development, Writing and Other Language Processes*, 3–50. London: Academic.

NEUMAN, SUSAN B., SERENE KOH, AND JULIE DWYER. 2008. Chello: The child/home environmental language and literacy observation. *Early Childhood Research Quarterly*, 23. 159–172.

NISHIMURA, BENSAKU, DAVID INGRAM, FRED C. C. PENG, AND PHILIP DALE. 1980. Emergence of functional articulation disorders. *Proceedings of the First International Congress for the Study of Child Language*, 439–447. Lanham, MD:

Univ. Press of America.

NISSEN, SHAWN L. AND ROBERT ALLEN FOX. 2005. Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *Journal of the Acoustical Society of America*, 118(4). 8.

NITTROUER, S., M. STUDDERT-KENNEDY, AND R. MCGOWAN. 1989. The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research*, 32. 12.

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

NITTROUER, SUSAN. 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.

NOWAK, M., PAWEL. 2006. The role of vowel transitions and frication noise in the perception of polish fricatives. *Journal of Phonetics*, 34. 139–152.

POOLE, I. 1934. Genetic development of articulation of consonants sounds in speech. *Elementary English Review*, 11. 159–161.

PRATHER, E., D. HEDRICK, AND D. KERN. 1975. Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders*, 53. 12.

R DEVELOPMENT CORE TEAM. 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`, ISBN 3-900051-07-0.

RINEY, TIMOTHY JAMES, NAOYUKI TAKAGI, KAORI OTA, AND YOKO UCHIDA. 2007. The intermediate degree of vot in japanese initial voiceless stops. *Journal of Phonetics*, 35(3). 439–443.

SCHELLINGER, K, SARAH. 2008. *The role of intermediate productions and listener expectations on the perception of children's speech.* Ph.D. thesis, University of Wisconsin.

SCOBBIE, E, JAMES, FIONA GIBBON, WILLIAM J. HARDCASTLE, AND PAUL FLETCHER. 2000. Covert contrast as a stage in the acquisition of phonetics and

phonology. Michael Broe and Janet Pierrehumbert, editors, *Papers in Laboratory Phonology V: language Acquisition and the Lexicon*, 194–203. Cambridge: Cambridge University Press.

SCOBBIE, J.M. 1998. Interactions between the acquisition of phonetics and phonology. K. Olson M. C. Gruber, D. Higgins and T. Wysocki, editors, *the 34th Annual Regional Meeting of the Chicago Linguistic Society*, volume II, 343–358. Chicago: Chicago Linguistics Society.

SHADLE, C. H. AND S. J. MAIR. 1996. Quantifying spectral characteristics of fricatives. *the International Conference on Spoken Language Processing (ICSLP 96)*, 1517–1520. Philadelphia, U.S.A.

SHADLE, CHRISTINE H. 1991. The effect of geometry on source mechanisms of fricative consonants. *Journal of Phonetics*, 19(3-4). 409–424.

SHADLE, CHRISTINE H. AND CELIA SCULLY. 1995. An articulatory-acoustic-aerodynamic analysis of /s/ in vcv sequences. *Journal of Phonetics*, 23. 53–66.

SMIT, A. B. 1990. The iowa articulation norms project and its nebraska replication. *Journal of Speech and Hearing Disorders*, 55(29-36).

SOLI, SIGFRID D. 1981. Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 70(4). 976–984.

STEVENS, KENNETH, ZHIQIANG LI, CHAO-YANG LEE, AND SAMUEL JAY KEYSER. 2004. A note on mandarin fricatives and enhancement. J. Cao G.Fant, H. Fujisaki and Y.Xu, editors, *From traditional phonology to modern speech processing.*, 393–403. Beijing: Foreign language teaching and research press.

STEVENS, K.N. 1998. *Acoustic Phonetics*. Cambridge: MIT Press.

STOEL-GAMMON, C., K. WILLIAMS, AND E. BUDER. 1994. Cross-language differences in phonological acquisition: Swedish and american /t/. *Phonetica*, 51. 146–158.

STONE, M, A FABER, L RAPHAEL, AND T SHAWKER. 1992. Cross-sectional shape and linguopalatal contact patterns in [s], [s], and [l]. *Journal of Phonetics*, 20. 253–270.

STUDDERT-KENNEDY, M. 1991. Language development from an evolutionary perspective. R. Schiefelbusch & M. Studdert-Kennedy N. Krasnegor, D. Rumbaugh, editor, *Language acquisition: Biological and behavioral determinants.* Hillsdale, NJ: Lawrence Erlbaum Associates.

TEMPLIN, M. 1957. *Certain Language Skills in Children.* Minneapolis: Univ. Minnesota.

TODA, M. 2005. Effect of palate shape on spectral characteristics of coronal fricatives.

TODA, M AND K HONDA. 2003. An mri-based cross-linguistic study of sibilant fricatives. *the 6th International Seminar on Speech Production.* Manly,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.

TYLER, A. A., G. R. FIGURSKI, AND T. LANGDALE. 1993. Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech and Hearing Research*, 36. 13.

WELLMAN, B., I. CASE, I. MENGERT, AND D. BRADBURY. 1931. Speech sounds of young children. *University of Iowa Studies in Child Welfare.*, 5. 1C82.

WHALEN, D. H. 1984. Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics*, 35. 15.

WHITE, DAVID. 2001. *Covert contrast, merger, and substitution in children's productions of /k/ and /t/.* Ph.D. thesis, The Ohio State University.

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

ZHU, HUA AND BARBARA DODD. 2000. The phonological acquisition of putonghua (modern standard chinese). *Journal of Child Language*, 27.