The Development of Phonation-type Contrasts in Plosives: Cross-linguistic Perspectives

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Eun Jong Kong, B.A., M.A.

*****

The Ohio State University
2009

Dissertation Committee:

Mary E. Beckman, Advisor

Elizabeth V. Hume

Cynthia Clopper

Approved by

_____

Graduate Program in Linguistics

ABSTRACT

Every spoken language has stops in its consonant inventory, and stop-vowel syllables such as [pa] and [ta] are among the first linguistic sounds to be identified in the babbling and first words of typically-developing children. A large majority of spoken languages also have at least two series of stops that contrast in their associated laryngeal gestures. This dissertation investigates how the acoustic details of the laryngeal contrast are related to the order in which children master the different stop phonation-type categories in their native languages. Analyses of a cross-sectional database of productions collected from Korean-, Japanese-, English- and Greek-acquiring children (2;0-5;11) supports some well-established claims about universal characteristics of children's stop phonation-type mastery patterns across languages, and also suggest the potential role of language-specific acoustic properties in explaining seemingly exceptional mastery patterns.

A starting point for this crosslinguistic comparison is Jakobson's (1941/1968) claim that there are implicational universals in the mastery of stop phonation types: the aspirated or voiced categories are mastered after the voiceless unaspirated category when a language has a contrast involving either aspiration or voicing. Using the acoustic measure of Voice Onset Time (VOT: the latency between oral constriction release and the onset of voicing), studies of many languages have shown that voiceless unaspirated stops (with a short lag VOT) are mastered before one year of

age, whereas voiceless aspirated stops (with a long lag VOT) are not mastered until two years of age in English (Macken and Barton, 1980a), Cantonese (Clumeck, Barton, Macken, and Huntington, 1981), and Thai (Gandour, Petty, Dardarananda, Dechongkit, and Mukongoen, 1986), and truly voiced stops (with a lead VOT) are mastered even later, at around age five in French (Allen, 1985), Thai (Gandour et al., 1986), Spanish (Macken and Barton, 1980b), and Hindi (Davis, 1995). Kewley-Port and Preston (1974) ascribed this universal order of mastery to the challenging aerodynamic and laryngeal control required to produce long lag VOT or lead VOT relative to short lag VOT.

I test this universal tendency of children's stop mastery and its relationship with the associated VOT properties in three relatively understudied languages (Greek, Japanese and Korean) and English. While the voiceless vs. voiced stops in English are successfully distinguished along the VOT range, the stop phonation-type contrasts in Greek, Japanese and Korean do not neatly fit into the three-way differentiation among truly voiced stops (with lead VOT), ordinary voiceless stops (with short lag VOT), and aspirated stops (with long lag VOT). We address three specific puzzles that arise from the VOT characteristics of stop categories in the target languages in relation to children's production accuracy, and examine the effects of various acoustic parameters in predicting the order of mastering stop categories.

The analysis tools include native speaker transcriptions of the stops produced by 100 children (2;0-5;11) acquiring Korean, Japanese, English or Greek to estimate the relative production proficiency, and the acoustic properties of children's stop productions measured with VOT, H1-H2 and *f0*. The stop productions collected from 20 adult speakers in each target language were examined to establish the objective norms of acoustic variables used in the native language. Based on the findings of these analyses, we ultimately aim to argue that the seemingly exceptional patterns

of mastering the speech sounds might be explained by knowledge of language-specific acoustic details of the categories. This would allow us to capture what can be stated as universal patterns in the mastery of stop phonation-types.

ACKNOWLEDGMENTS

During my seven long years of stay in the program, I am indebted to many people in the department. Without them, this dissertation would not have been possible.

First of all, I would like to express my enormous appreciation to my advisor, Dr. Mary E. Beckman, for her support and encouragement in every step of this dissertation and also throughout my doctorate program at the Ohio State University. She explicitly taught me what phonetic science is (as well as how to make the 'real salad') and also implicitly taught me what a scientist should be and how exciting it is to be a phonetician. I am deeply grateful for her guidance and care for me for all those years, and feel fortunate to have her as a role model in my academic career.

I also would like to thank Dr. Jan Edwards, at University of Wisconsin-Madison, for her advice and encouragement for this project. The Paidologos project introduced me to the field of language acquisition, motivated me to pursue this as the dissertation topic, and practically made it possible to happen. I thank both Mary and Jan for providing me with such a helpful and friendly environment of studying child language acquisition.

My appreciation goes to my dissertation committee members, Dr. Elizabeth V. Hume and Dr. Cynthia G. Clopper. They have given me insightful comments on this project, which were from all different perspectives of understanding 'sounds'. I sincerely appreciate their encouragement and interest in my research.

There are many people in Seoul that I owe many thanks to as well. Dr. Keeho Kim, my M.A. thesis adviser at Korea University, and Dr. Mirah Oh have been mentoring me, so that I can survive the doctorate program at OSU.

Last but not least, I would like to express my dearest appreciation to my beloved family: my parents, sisters and aunts who trust and love me unconditionally.

VITA

1999 ..... B.A., English Language and Literature,
Korea University

2002 ..... M.A., English Language and Literature,
Korea University

2002 - 2003, 2005 - 2008 ..... Research Assistant,
The Ohio State University

2003 - 2005, 2008 - 2009 ..... Teaching Assistant,
The Ohio State University

PUBLICATIONS

1. Kong, Eun Jong (2004). The role of pitch range variation in discourse structure and intonation structure of Korean. In *Proceedings of the International Conference on Spoken Language Processing*. Jeju, Korea.

1. Kong, Eun Jong, Beckman, Mary E., and Edwards, Jan. (2007). Fine-grained phonetics and acquisition of Greek voiced stops. In *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbruken, Germany.

FIELDS OF STUDY

Major Field: Linguistics Specialization: Phonetics

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Every spoken language has stops in its consonant inventory, and stop-vowel syllables such as [pa] and [ta] are among the first linguistic sounds to be identified in the babbling and first words of typically-developing children. A large majority of spoken languages also have at least two series of stops that contrast in their associated laryngeal gestures. For example, Spanish has voiced stops [b, d, g] that contrast with voiceless stops [p, t, k], and Cantonese has aspirated stops [p^h, t^h, k^h] that contrast with unaspirated [p, t, k]. These contrasts in pre-vocalic position are typically characterized in terms of differences in Voice Onset Time (VOT), a measure of the latency between the stop's oral release and the onset of periodic energy (voicing). Voiced stops have "lead" VOT (with voicing starting well before the release), whereas aspirated stops have "long lag" VOT (with voicing starting well after the release). Voiceless unaspirated stops then are a kind of default, with voicing starting simultaneous with or at a very "short lag" after the release.

This dissertation is about the acquisition of stop phonation-type contrasts across languages. It discusses the role of acoustic properties such as VOT in describing stops that contrast in their laryngeal properties in different languages, and in predicting the pattern of mastery. That is, we investigate whether the order in which children master the different stop phonation types that contrast in their native languages is related to the acoustic details of the categories. We specifically

focus on contrasts that do not neatly fit into the three-way differentiation among fully voiced stops (with lead VOT), ordinary voiceless stops (with short lag VOT), and aspirated stops (with long lag VOT). Using a cross-sectional database of productions collected from Korean-, Japanese-, English- and Greek-acquiring children aged 2 years through 5 years, we address three puzzles regarding the interaction between the acoustic properties of the categories in contrast, and the children's production proficiency. These puzzles arise in the context of the well-demonstrated utility of VOT in comparing stop phonation-type contrasts across languages, and in explaining commonly attested mastery patterns.

Since the seminal study by Lisker and Abramson (1964), VOT has been used as a successful acoustic measure in comparing the stop categories across many languages. VOT is a continuous measure of the temporal relationship between two acoustic events that signal the onset of vocal fold vibration and the oral constriction release, but the three qualitatively different relationships of (1) "before" versus (2) "simultaneous with" versus (3) "well after" does capture the three most commonly attested categories across languages. For example, in Lisker and Abramson's original crosslinguistic investigation of stop VOT distributions across eleven languages, the two stop phonation types in Dutch, Hungarian, Spanish, and Tamil could be classified as "true" voicing contrasts that differentiated between lead and short lag values, whereas the two stop phonation types of Cantonese and English could be classified as aspiration contrasts between the long lag VOT range for aspirated stops and the short lag VOT range for unaspirated stops. The three stop phonation types of Eastern Armenian and Thai were a simple union of these two two-way contrasts. Only the three-way contrast in Korean and the four-way contrasts in Hindi and Marathi could not be captured by these three categories of lead versus short lag versus long lag.

This three-way distinction among lead, short-lag, and long-lag ranges for VOT as the relevant values for describing voicing and aspiration contrasts has also been useful in investigating phonological development. That is, a convenient way to talk about children's mastery of their native-language categories is to see what VOT values are produced by young children in their first words, and at what age children's VOT values begin to look like adult norms for the target types. For example, Clumeck et al. (1981) show that Cantonese-speaking children's earliest stops are all realized with short lag VOT values. That is, the youngest children essentially substitute the "default" unaspirated type for the aspirated stops, and aspirated stops with long lag VOT only appear around 24 months. Gandour et al. (1986) show that 3-year-old Thai-speaking children already produce aspirated stops with long-lag VOT values, but voiced stops with lead VOT are characteristic of only older children's productions. Younger children produce only short-lag values, effectively substituting voiceless stops for voiced ones. Researchers such as Kewley-Port and Preston (1974) relate these mastery patterns to the more precise motor control that is required to meet the challenging aerodynamic conditions for voicing lead and long lag.

In this dissertation, we use VOT and other acoustic properties to compare the mastery patterns for stop phonation types across four target languages: English, Korean, Japanese, and Greek. Mastery of the English contrast has been studied extensively, and the mastery pattern is very much like that for Cantonese and other languages that have an aspiration contrast. The other three languages, however, pose various puzzles. First, as Lisker and Abramson (1964) showed, Korean is anomalous in having a three-way contrast that does not use the "default" voiceless category, but instead contrasts "tense" laryngealized stops with two types of more or less long-lag stops. Japanese is anomalous because the voiced stop has a short lag variant with VOT values that overlap with those of the voiceless stop, which is "mildly aspirated",

with VOT values intermediate between the two Cantonese or English categories. The Greek pattern is puzzling in that Greek children master voiced stops at a much earlier age than Dutch-, Thai-, French-, and Spanish-learning children. In the rest of this chapter, we review the previous literature in more detail to elaborate on this description of the Korean, Japanese, and Greek puzzles, and to motive our own choice of other acoustic cues to examine for the Korean and Japanese contrast.

1.1 Stop laryngeal contrasts and Voice Onset Time

1.1.1 VOT and phonological development

The division of the VOT continuum into the three qualitatively different ranges originally proposed by Lisker and Abramson (1964) has been adapted as useful method of analysis in assessing which stop category in a laryngeal contrast is mastered by children before others. This objective way of describing the adult-like realization of the stop category has made it possible to capture two consistent trends in children's speech development of laryngeal contrasts across languages and to generalize them as a well-motivated developmental universal.

One common finding in the mastery of stop consonants across languages is that the voiceless unaspirated stop is mastered before any other contrastive stops, which is demonstrated by the short lag VOT value in children's stop productions as early as in canonical babbling at 6 - 8 months. Jakobson (1968) stated this mastery order as the implicational universal that "so long as stops in child language are not split according to the behavior of the glottis, they are generally produced as voiceless and unaspirated" (p.14), which predicts that aspirated or voiced categories are mastered after the voiceless unaspirated category when a language has a contrast involving

either aspiration or voicing. As demonstrated in studies of various languages (Table 1.1), it is the short lag VOT that typically serves as an undifferentiated acoustic form of children's early stops no matter which laryngeal contrasts the language has in its stop consonants.

Kent (1981) hypothesized that infants may have a propensity to a certain phonetic category that is shaped by the non-linear nature of articulatory-acoustic parameters experienced through their babbling based on Quantal Theory. Stevens's (1972) Quantal Theory suggests that the relation between articulation and sound output is not linear in that there are articulatory conditions where a small change in the acoustic parameter results in a large change in the acoustic properties and there are other conditions where a large change in articulation yields insignificant change in the acoustic characteristics. Stevens (1972) uses these non-linearities to explain distinctive features: each discrete unit of linguistic information would potentially be realized at a region where the acoustic property is generated without precise positioning of the articulatory gestures. Kent (1981) related this idea to the mastery of phonetic categories. He suggested that children explore the acoustic consequences of a wide variety of articulatory movements while babbling and find stable acoustic regions where the changes in articulatory gestures do not result in changes in the acoustics. Children's mastery of common speech categories would exhibit language-universal characteristics by being affected by these physical constraints.

Kewley-Port and Preston (1974) ascribed the preponderance of short lag VOT values in children's early stops to the lesser articulatory requirements on producing this outcome. Tracing the development of apical stops with respect to VOT in English-acquiring children, from the children's earliest stops in canonical babbling at 6 months, they found randomly distributed VOT values at first, ranging from a few

tokens with lead VOT to a few with long lag VOT. In 9-10 months, however, the children's VOT values were concentrated at the short lag range, regardless of the target stop voicing contrast. The production of short lag VOT is relatively easy in that it is achieved by the glottis opening at any time during the oral occlusion. Moreover, it can even be achieved without opening the glottis, if other maneuvers for relieving intra oral air pressure are not performed. Before mastering the precise temporal control between the gestures of the glottal opening and the oral constriction release for long lag VOT or lead VOT, children might be making imprecise coordinations between the articulatory gestures producing the acoustic consequence of short lag VOT value. The crosslinguistic findings of universally early mastery of the voiceless unaspirated stop would be predicted by its definition in terms of short lag VOT, which is the value that has the least demanding articulatory control.

Another common crosslinguistic trend in the development of laryngeal contrastive stops is that the aspirated stops are mastered at about age two and the voiced stops are mastered at about age five as summarized in Table 1.1. In studies documenting the VOT characteristics of stop consonants produced by children, the long lag VOT category emerges in children's productions at about age two as a distinct VOT cluster in a distribution, while lead VOT is not adult-like in children's productions even until four or five. These studies of children's VOT realizations across languages have described the aspirated stop as a category mastered later than the unaspirated stop in English (Macken and Barton, 1980a), Cantonese (Clumeck et al., 1981), Thai (Gandour et al., 1986) and Hindi (Davis, 1995), and the voiced stop as a category mastered even later in French (Allen, 1985), Thai (Gandour et al., 1986), Spanish (Macken and Barton, 1980b) and Hindi (Davis, 1995). For example, Gandour et al. (1986) described the time course of VOT changes in order to define the mastery order among three-way contrastive Thai stops such as unaspirated, aspirated

and voiced stops. They found that there was no overlapping VOT range between a long lag VOT for the aspirated stop and a short lag VOT for the unaspirated stops in three year old Thai-acquiring children's productions, whereas VOT values for the voiced stop were not adult-like until the age of five. The relatively late mastery of lead VOT for the voiced stop was true in languages such as French and Spanish where the stop cognates are contrastive between voiced and voiceless stops.

The relative difficulty in achieving the target acoustic outcomes of long lag VOT for aspirated stops and lead VOT for voiced stops can predict the relative mastery order among the stop categories (Kewley-Port and Preston, 1974). While the production of short lag VOT is achieved by the glottis opening at any time during the oral occlusion, the production of long lag VOT requires a precise temporal adjustment between the oral articulation and the glottal opening gesture. The temporal coordination needs to be precise enough to align the maximum glottis with the oral release. The capability of temporal adjustments would take time to master, yielding the adult like long lag VOT at a later age than the short lag VOT.

In the same vein, the delay of the mastery of voiced stops (particularly in initial position) is predicted by the challenging motoric demand in realizing adult-like lead VOT. The production of lead VOT requires that the glottal adduction gesture be made prior to the oral constriction release and also that supraglottal air pressure be sufficiently lower than the subglottal pressure for vocal fold vibration to begin during the closure (Westbury and Parris, 1970; Westbury, 1983; Westbury and Keating, 1986; Keating, 1983). Under the condition where one end of the vocal cavity is closed, it is not easy to maintain a supraglottal pressure lower than the subglottal pressure. The pressure differential can be achieved by lowering the larynx or performing other maneuvers that effectively enlarge the oral cavity (Catford, 1977; Ohala, 1983).

Language (source)	Age and VOT pattern
	asp. vs. unasp.
English (Kewley-Port and Preston, 1974)	Separate VOT peak for /t^h/ at 75 weeks
English (Macken and Barton, 1980a)	Longer VOT for aspirated stops beginning at 1;5-1;7
Cantonese (Clumeck et al., 1981)	Separate VOT peak for aspirated stops at 2;4
	voiced vs. voiceless
Spanish (Macken and Barton, 1980b)	Instantiations of lead VOT for voiced stops rare until 4;0. Consistent use of spirantization instead of lead VOT for voiced stops.
French (Allen, 1985)	Rare occurrence of voiced stops in the data (1;9-2;8). Strong tendency of having voiced segment preceded to the voiced stop target.
	voiced vs. voiceless vs. asp.
Thai (Gandour et al., 1986)	Significantly different mean VOTs between aspirated vs. unaspirated at three. No adult-like lead VOT values for voiced stops until 5;0.
Taiwanese (Pan, 1994)	Clear long lag VOT for aspirated stops at 28 months. Lead VOT voiced stops at 34 months
	vd asp. vs. vl asp. vs. vd unasp. vs. vl unasp.
Hindi (Davis, 1995)	No clear difference between voiced asp. and voiceless asp. even in 6-year-olds' production

Table 1.1: Summary of earlier studies

Children adapt various strategies to fulfill the aerodynamic requirements for lead VOT. For example, Macken and Barton (1980b) found that it was not until age four that Spanish-speaking children produce adult-like voiced stops. Before that age, the target voiced stops in initial position are frequently spirantized (just as they are intervocalically in adult speech). The spirantized variants can be interpreted as resulting from a loose seal during the oral constriction interval. That is, children might adapt the intervocalic allophone to lower the supraglottal pressure in word-initial position as well. Similarly, Allen (1985) found only a few measurable lead VOT tokens for voiced stop targets in French-speaking children's word-initial productions (1;9-2;8). Many tokens for voiced stop targets were not measurable in terms of VOT because they were preceded by other segments. Interestingly, the preceding segments were voiced (mostly nasals). This can be interpreted as an intention to take advantage of voicing of the preceding voiced segment in making a lead VOT voiced stop. Especially, the preceding nasal segments could have resulted from the children's use of nasal cavity opening as a way to meet the supraglottal pressure for the vocal fold vibrations. Along the same line, Whalen, Levitt, and Goldstein (2007) observe that most prevoiced stops produced by French-speaking children in the late babbling stage are not adult-like in that they show energy at high frequencies in addition to the typical low frequency energy for the voice bar. They also show a weak burst when the oral constriction is released, suggesting that intraoral air is vented through the open nasal passage. These findings of other language-specific acoustic variants for voiced stops before the emergence of adult-like lead VOT patterns reflect the extreme motoric demands for producing lead VOT in word-initial position. The delay of mastery of voiced stops and aspirated stops is predicted by this relative difficulty in controlling the articulation for the corresponding acoustic output of VOT.

1.2 Some remaining questions

While VOT is useful tool to describe the pattern of mastery of stop cognates in word-initial position, there still remain some questions regarding how to describe the mastery of stop categories whose laryngeal contrast VOT cannot neatly differentiate.

This section describes three specific questions about the acoustic characteristics of stop phonation types in some languages that do not fit neatly into the types described above and the difficulties that these 'non-canonical' types pose for predicting relative order of mastery. First of all, we ask how to predict the order of mastery of stop categories in a language with a phonation-type contrast that uses a feature other than voicing and aspiration. For example, what can we predict about Hindi breathy-voice stops or Korean tense stops? Secondly, we ask how to predict the mastery of stop categories when one of the categories use intermediate values of VOT. For example, what can we predict about the voicing contrast in Quebec French, which has many short lag values for /b, d, g/and values between short and long lag for /p, t, k/. Thirdly, we ask what additional acoustic characteristics we can measure to begin to explain differences across languages in children's mastery of prevoicing lead. For example, the Taiwanese-acquiring children studied by Pan (1994) differentiated their voiced stops from voiceless unaspirated stops at about 32 months, well before the European French-acquiring children described in Allen (1985).

1.2.1 Voiced aspirate and tense stops

The first question concerns how the pattern of mastery of the stop phonation-type contrast would be described when one of the stop categories is not defined as voiced, voiceless unaspirated, or aspirated. As Lisker and Abramson (1964) pointed out, there are stop types such as Hindi voiced aspirated stops and Korean tense stops that

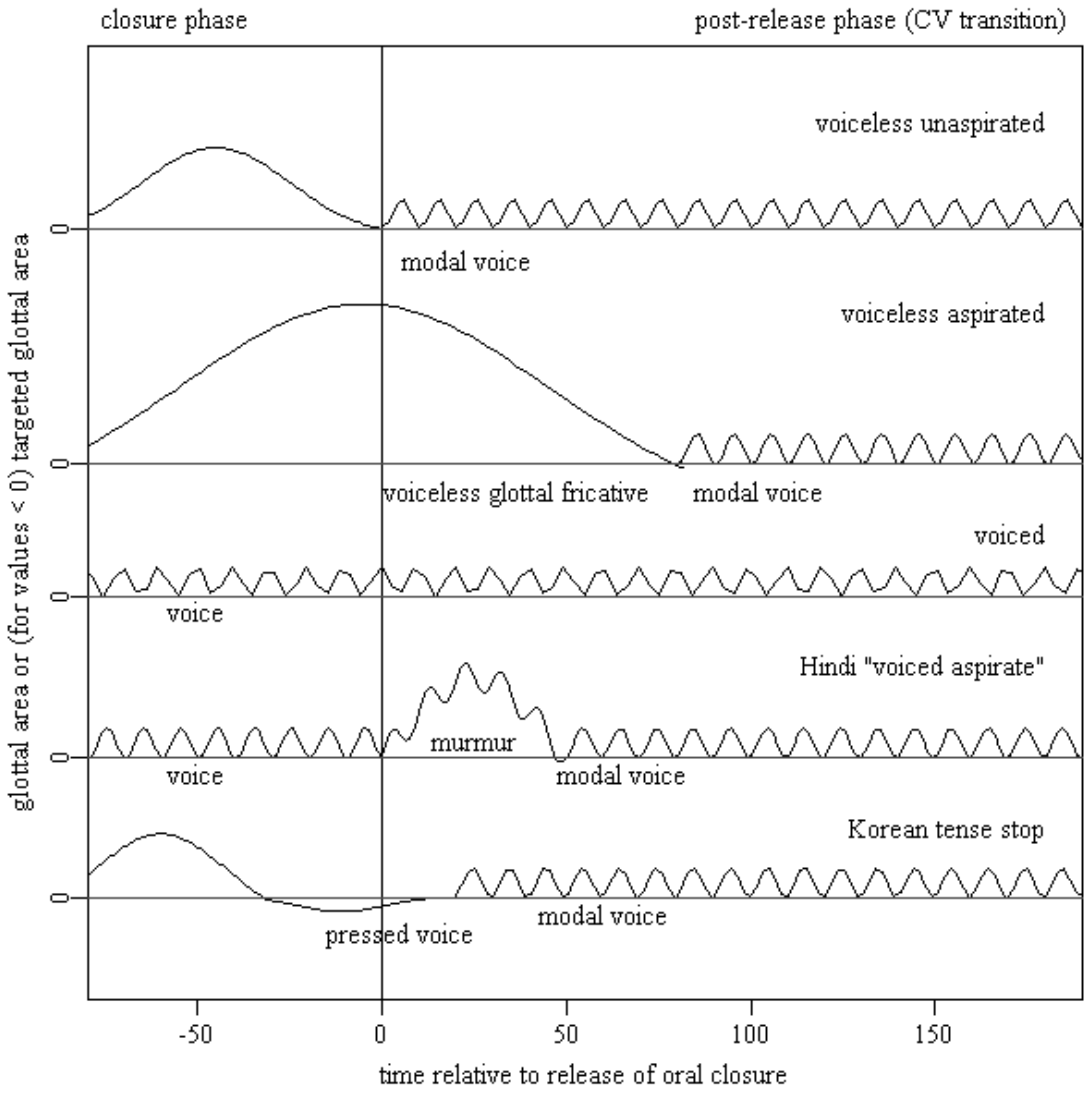

Figure 1.1: Schematic illustration of glottal area over time in productions of plosives with different phonation types.

do not belong to any of the three common types of stops, resulting in overlapped VOT values with at least one other contrastive stop category in the language.

In plosive consonants, phonation types are defined articulatorily by the glottis condition over time from the oral closure to an interval after the constriction release,

as shown in Figure 1.1. While glottal adduction and the presence of vocal fold vibration during the closure differentiates the 'voiced' and 'voiced aspirate' categories from others, maximal glottal opening at or after the release of the oral constriction differentiates 'voiceless' and 'voiceless aspirate' categories from the rest of the panels. If we think of the glottal adduction and accompanying maneuvers to expand the supraglottal cavity as a composite 'voicing gesture' and the glottal abduction at or after release as an 'unaspirated gesture', then 'voiceless unaspirated stop' is the default or 'unmarked' category that lacks both gestures and the 'voiced aspirate' is the maximally marked category that carries both properties. However, the 'Korean tense stop' cannot be described in terms of either a voicing gesture or an aspiration gesture.

When the phonation types are to be described in acoustic space, VOT is not a sufficient domain to characterize the 'Hindi voiced aspirate' nor the 'Korean tense' stop. As schematized in Figure 1.1, the voiced aspirate has a larger glottal opening than the voiced unaspirate after the oral constriction release, which ceases vocal fold vibrations for a short duration after burst until the glottis is set narrow enough to resume the vocal fold vibration. The systematic acoustic difference between the aspirated voiced stop and the unaspirated voiced stop is found after the release where the modal voicing is delayed after the burst release. Davis (1994) took the lag time into consideration in her acoustic characterization of voiced aspirated stop separated from voiced unaspirated stop in Hindi, demonstrating that the lag between the burst and the onset of the second formant in the following vowel successfully captured the difference between the aspirated voiced stop and the unaspirated voiced stop.

Similarly, the acoustic characteristics of Korean tense stops are not fully described by VOT. The glottal condition of the Korean tense category does not pattern with that of the voiced stop prior to the release nor with that of the aspirated stop

after the release. What characterizes the tense stop is a laryngeal muscle tenseness that suppresses the vocal fold vibration before the oral constriction release and delays the voicing onset until immediately after the release (Kagaya, 1974; Hardcastle, 1973; Hirose, Lee, and Ushijima, 1974; Dart, 1987).

According to the electromyographic study in Hirose et al. (1974) where they examined the role of different muscle activities in the productions of the three stop phonation-types in Korean, tense stops are characterized as having an increasing activity of vocalis and lateral cricoarytenoid before the oral constriction release. This results in an increase in inner tension of the vocal folds as well as glottal constriction immediately after the release. The aspirated stop is distinguished from the lax and tense stop categories by having an adduction muscle activity, which is suppressed during the closure and becomes active quickly at the release. Lax stops lacked the described positive laryngeal muscle activities to be contrastive with the other stops. These characteristics of lax stop produce the passive voicing of intervocalic lax stops in Korean.

Dart (1987) measured the intra-oral pressure during the occlusion and the peak air flow after articulatory release using a respiratory mask and a pressure transducer. The findings were that tense stops tend to have a higher intra-oral pressure and a lower airflow than lax stops after the release. This aerodynamic model had to postulate a tenser vocal tract wall in order to make the oral cavity pressure higher, the air flow lower, and the glottal opening smaller. Hardcastle (1973) speculated from the aerodynamic and acoustic data that there is physiological adjustment in tense stops such as stiffened vocal cords and pharynx. The vocal cord tension was assumed to exist in the production of tense stops because the vocal cord vibrations did not begin as soon as aerodynamic conditions such as supraglottal pressure drop and the complete glottal adduction was met. It has been observed that the initiation

of voicing in tense stops is made by lowering the larynx in order to enlarge the vocal cavity to allow desirable conditions for vocal fold vibration (Kagaya, 1974).

The muscle tenseness of tense stop affects not only VOT but also other acoustic dimensions such as fundamental frequency or creaky voice quality. The VOT of the Korean tense stop is in the short lag region. In addition to having this short lag VOT, it is further characterized as having a higher *f0* immediately after voice onset due to the physical tension in the vocal cords, and a somewhat creaky voice quality, due to the vocal cords being pressed (Kagaya, 1974; Hardcastle, 1973; Hirose et al., 1974; Dart, 1987; Jun, 1993). This segmental property of having a higher *f0* is phonologized in the intonation structure of Seoul Korean, so that the tonal pattern at the beginning of the accentual phrase is determined by the phonation-type of the phrase-initial consonant segment (Jun, 1993, 1998). The tense or aspirated consonant phonation-type is associated with a H tone-initial accentual phrase (HHLH) while the lax phonation-type makes for a L tone initial accentual phrase (LHLH). Thus, VOT is only one of several acoustic characteristics that differentiate the tense stops from the lax and aspirated stops.

In assessing how accurately children realize these kinds of 'non-canonical' stop phonation-type categories in the course of phonological development, children's productions need to be examined with respect to all of the relevant acoustic dimensions, not just VOT.

1.2.2 Intermediate VOT range

The second question is related to the existence of languages with an aspiration contrast that have a categories with VOT values that are intermediate between the short lag category and the long lag category. While this type, with intermediate lag VOT, had not been originally documented in Lisker and Abramson (1964), it is reported

in later studies of VOT distribution in languages other than the eleven target languages investigated in Lisker and Abramson (1964). The existence of this category with intermediate values of VOT has, thus far, always involved a contrast with another category that has at least some short lag values, creating an overlap between two lag VOT ranges. The overlap suggests that the VOT dimension alone cannot provide sufficient acoustic information to distinguish homorganic stops that are in contrast in the language. This, in describing the mastery of the stop categories, in these languages, then, we must consider other acoustic dimensions to complement the ambiguous acoustic cues from the overlapping VOT values.

The VOT mean comparison of voiceless stops across 18 languages in Cho and Ladefoged (1999) provided the evidence that the VOT range can vary beyond what Lisker and Abramson (1964) suggested earlier. There was a group of languages whose voiceless stop VOT means were between two peak values of VOT distributions of the typical short and long lag ranges. While this ambiguous VOT mean value might have been obtained as an artifact of having target languages of various types of phonation-type contrast, such as no contrast in Banawá and a three-way contrast among unaspirated, glottalized and aspirated stops in Hupa, there is also evidence of an intermediate VOT range in languages such as Canadian French, Hebrew and Japanese that are known to be contrastive between voiced and voiceless stops.

Caramazza, Yeni-Komshian, Zurif, and Carbone (1973) found that the VOT range of voiceless stops by monolingual Canadian French speakers was between the short lag and the long lag ranges used by English monolingual speakers. This intermediate VOT value produced a substantial overlap between the VOT ranges of voiced and voiceless stops in Canadian French. In perceiving /t/ tokens whose VOT values were systematically manipulated to vary in every 5ms or 10ms steps, Canadian

French listeners were uncertain about the judgment of voiceless stop /t/ with overlapping VOT values, suggesting that VOT is not a sufficient acoustic cue to the voicing distinction in Canadian French, unlike English. Caramazza and Yeni-Komshian (1974) followed up by comparing Canadian French stop VOT distributions with the stop VOT distributions produced European French. They found that voiceless stop VOT values in Canadian French were longer than those in French. In addition, the voiced stops in Canadian French are realized not only with lead VOT values but also sometimes with short lag VOT values, whereas the voiced stops in European French consistently have lead VOT values. Thus Canadian French has overlapping VOT values between the short lag voiced stops and the intermediate lag voiceless stops.

Sundara (2005) showed that the spectral characteristics of the stop burst play a role in disambiguating the voiced stop from the voiceless stop, despite the overlapping VOT in Canadian French. For example, the burst intensity relativized by the intensity of the following vowel was significantly higher in voiceless stops than in voiced stops in Canadian French. The same spectral properties in Canadian English failed to make a statistically significant difference between aspirated and unaspirated stops whose VOT ranges do not overlap.

Hebrew and Japanese are two more examples where the voiceless stop is produced with intermediate VOT values (Raphael, Tobin, Faber, Kolier, and Milstein, 1995; Riney, Takagi, Otaa, and Uchida, 2007). Raphael et al. (1995) investigated the VOT distributions of stops spoken by Hebrew-speaking adults and children (10;0-11;0) in Israel and in USA, finding that the voiceless stops in Hebrew have intermediate VOT values. In Japanese, this intermediate VOT range of the voiceless stop in Riney et al. (2007) reflects the change of voiceless stop VOT since the VOT measurement of Japanese voiced vs. voiceless stops in Homma (1980). While the voiceless stop VOT in Homma (1980) ranged at a short lag range, the values measured later in Riney

et al. (2007) tend to have a longer VOT value intermediate between a short and a long lag VOT. Similar to the VOT overlap between voiceless and voiced stops in Canadian French, Japanese voiced and voiceless stops are not sufficiently separated from each other due to this intermediate VOT of voiceless stops and the short lag VOT variants of voiced stops in Japanese (Takada, 2004b).

The existence of an intermediate VOT range suggests that the two-way voicing contrast between voiced and voiceless stops, one of the most common two-way stop phonation-type contrasts across languages, is not necessarily described as a category with lead versus a category with short lag values along the VOT continuum. The VOT ranges of purportedly identical stop voicing categories can be language-specific. The resulting overlap between the intermediate VOT range and the short lag VOT range for contrastive stop categories implies the potential role of other acoustic dimensions as a complementary cue to VOT in distinguishing one category from the other. For an accurate characterization of the mastery pattern of stop voicing categories in languages with such an 'intermediate lag' type, an additional acoustic dimension besides VOT needs to be examined.

1.2.3 Prevoicing in voiced stop

Finally, the third question involves the spectral characteristics of the maneuvers that languages use to make 'true' voiced stops that the temporal measure of VOT cannot capture in nature. Depending on the degree of oral seal tension and the availability of the nasal air-passage, the spectral quality of prevoicing varies in many different ways.

Voiced stops can be spirantized as in the allophonic variant of Spanish intervocalic voiced stops (Lewis, 2002; Romero, 1995; Cole, Hualde, and Iskarous, 1999). The spirantized voiced stop is made with a less intense oral seal during the stop closure,

which allows a leak of oral airstream. This articulatorily reduced form of a voiced stop would make it conducive to maintain the transglottal air-pressure desired for the vocal fold vibrations. Acoustically, the spirantized allophone in the word-medial position contains a certain degree of high frequency non-periodic energy, although it was too little to be defined as a fricative (Romero, 1995). In Cole et al. (1999), the higher degree of spirantization in Spanish intervocalic voiced stops was described as a higher relative energy of the consonantal duration.

Purely voiced stops would have modal prevoicing indicated by periodic low frequency energy given that the source is restricted to the glottal air stream with the oral air passage tightly sealed. One possible kind of variant of a voiced stop is a prenasalized prevoiced stop which is made by having a velopharyngeal port opening before the release of the oral closure. This would make it easier to initiate the prevoicing of a voiced stop since the substantial opening in the airway through the velopharyngeal port makes the aerodynamic condition for vocal fold vibration met by lowering the supraglottal pressure during the oral occlusion. Some languages have this type of prenasalized variant as a category in contrast with purely voiced stops. For example, there are prenasalized stops in contrast with stops in Moru (Burton et al., 1992) and in Fijian (Maddieson, 1989). Other languages have this prenasalized variant only at the parametric level as in Greek (Arvaniti and Joseph, 1999; Okalidow, Petinou, Theodorou, and Karasimou, 2002).

Burton et al. (1992) examined the temporal and spectral characteristics of voiced stops, prenasalized stops and nasals in Moru where all three types of segments are phonologically contrastive. While the temporal characteristics of VOT failed to distinguish voiced stops from prenasalized stops, the spectral characteristics of the initial sonorant interval could differentiate the prenasalized stops from the purely voiced stops. This raises the need to examine the spectral characteristics of voicing

lead apart from VOT to make a correct characterization of what children should master in the ambient language.

The particular voiced stop variants available in a language with a 'true' voicing contrast appears to influence children's mastery pattern of voiced stops. The Spanish-learning children in Macken and Barton (1980b) consistently substituted the spirantized allophones of voiced stops for their word-initial voiced stops, suggesting that children adapt the strategy of having a continuant airway to overcome the difficult aerodynamics of making a purely voiced stop.

While VOT still plays a sufficient role in distinguishing the prevoiced stop with negative VOT from the voiceless stops with positive VOT, the subtle phonetic details during the stop closure would not be captured in the temporal domain. More importantly, the suggested variants of prenasalized stops or spirantized allophones of voiced stops have been pointed out as substituted acoustic outcomes in young children's intended voiced stops. The examination of fine-grained acoustic details in the voiced stop is necessary to assess these variants of children's voiced stop productions.

1.3 Other acoustic parameters

While VOT has been demonstrated as a reliable acoustic cue to the stop laryngeal contrast, the consonant voicing/aspiration contrast is conveyed by several other acoustic cues as consequences of glottal gestures and aerodynamic conditions at the glottis.

1.3.1 First formant: F1

Associated with the aperture between the articulator and the ceiling of the mouth, the transition of the first formant (F1) is recorded in the spectral signal differently depending on how long the aperiodic source of energy continues before the periodic

energy appears. The voiceless aspirated stops lack trace of the F1 transition in the spectral record because the articulator for the oral constriction (e.g., the tongue tip for ab alveolar stop) is likely to finish moving to the configuration for the following vowel before the excitation by aperiodic source of energy begins. In contrast, the unaspirated stops have a longer spectral record of the F1 transition because the articulator is likely to be in motion when vocal cord vibration starts to excite (Stevens and Klatt, 1974; Summerfield and Haggard, 1977; Summerfield, 1982). As a result, the duration of the F1 transition for unaspirated stops will be shorter than that for aspirated stops, covarying with VOT (Park, 2002). Also, this 'cut-back' of the F1 transition in unaspirated stops makes the onset F1 value at the voicing onset different from the onset F1 value after aspirated stops.

The F1 cues English listeners in judging the voicing category of stop consonants. Stevens and Klatt (1974) demonstrated that English listeners tend to reset the phonemic boundary for the voiced stops in the VOT dimension accommodating the F1 characteristics at the vocalic onset. When there was a slow and long transition of F1, the listeners judged the stimuli with a relatively long VOT as an aspirated stop. Summerfield (1982) varied the F1 onset frequency instead of duration of transition to examine whether the phoneme boundary along VOT continuum shifts according to the changes of F1 onset frequency. He found that the VOT boundary was shortened as the F1 onset frequency increased. Similarly, Benkí (2001) found that subjects responded to each VOT-varying stimulus with more [-voiced] bias when the stimulus had a higher F1 onset frequency with a shorter transition duration than when it had a lower F1 onset frequency with a longer transition duration. A similar effect of F1 was found in Icelandic listeners' identification patterns of consonant voicing contrast (Pind, 1999), and in Korean speakers' production patterns of stops (Park, 2002).

1.3.2 Fundamental frequency: *f0*

f0 cues the consonant phonation-type across languages. It captures an artifact of the laryngeal and oral motoric mechanism in consonant phonation which is primarily affected by tensions in cricothyroid muscles and thyroarytenoid muscles with an interaction with subglottal pressure (Honda, 1995; Titze, 1995; Hombert, Ohala, and Ewan, 1979; Stevens, 2000). The extrinsic laryngeal muscle activities related with the tongue and vertical larynx movement also affect *f0* values after the consonant (Honda, 1995; Kim, Honda, and Maeda, 2005). Speakers of various languages manipulate this acoustic properties of *f0* by linguistically encoding the consonant phonation-types.

Ohde (1984) showed that English voiceless stop /p/ ([p^h] word-initially) had a higher *f0* perturbation at the vocalic onset than /b/ did, when the *f0* was measured over the first five glottal pulses after the vocalic onset. Haggard, Summerfield, and Roberts (1981) also found the same tendency that English voiceless stops have a higher *f0* than voiced stops. They conducted perception experiments where the two acoustic parameters of *f0* and VOT were manipulated to vary across the stimuli in order to investigate the influence of *f0* on the stop voicing category identification. The results showed that the phoneme boundary between unaspirated stops and aspirated stops shifted sensitively to the *f0* onset variable. Similarly, Whalen, Abramson, Lisker, and Mody (1993) found that a fast response time was obtained when listeners selected /d/ with a relatively lower *f0* and /t/ with a relatively higher *f0*. This facilitating effect of *f0* in making a decision of stop voicing category was observed even when the VOT values of the stimuli clearly cued the stop voicing categories.

Shimizu (1989) examined five Asian languages to demonstrate the use of the *f0* parameter in distinguishing the stop phonation type contrast. The *f0* was considerably higher at the vowel onset after voiceless than voiced stops in Japanese, Burmese,

Thai, Hindi and Korean (higher *f0* in tense than in lax). In Korean, the *f0* values of tense and aspirated stops are higher than those of lax stops, and are phonologized in the intonational phonology of the Seoul dialect (Jun, 1993). This critical role of *f0* in the distinction among three Korean stops is discussed in detail in Chapter 3.

The segmental properties of *f0* also exist in other tone languages such as Cantonese (Francis, Ciocca, Wong, and Chan, 2006) and Yoruba (Hombert, 1977). Cantonese aspirated stops have a higher *f0* at the following vowel onset shortly before the tone is realized in the production. The *f0* properties cued the stop aspiration contrast in Cantonese listeners' perceptual pattern when they were asked to select the word with the tone that varied *f0* at vocalic onset. The listeners' correct responses to the aspirated stop were significantly affected by the vocalic onset *f0*.

1.3.3 Voice quality: spectral tilt

The aspiration or voicing characteristics of plosive consonants influence the voice quality of the adjacent vowels due to the glottal configuration set for the phonation. Breathy voice is made by a posterior glottal opening posture during the glottal closing or by a non-simultaneous glottal closing. It is associated with a bigger percentage of the glottal cycle during which the glottis is open (the increased open quotient) and less abrupt glottal closing gestures (Holmberg, Hillman, and Perkell, 1988; Hanson, 1997). Acoustic correlates of physiological configurations for breathy phonation are the first harmonic amplitude rise, and reduces the amplitude of higher frequency (Holmberg et al., 1988; Klatt and Klatt, 1990; Hanson, 1997; Hanson and Chuang, 1999; Gordon and Ladefoged, 2002).

Voice quality is quantified by measuring the spectral tilt of the amplitude of the first harmonic(H1) in relation to the amplitude of higher frequency such as the amplitude of the second harmonic(H2), the first formant amplitude(A1) or the third

formant amplitude(A3). The breathy voice quality of an increased open quotient is correlated with H1-H2 (Holmberg et al., 1988). A greater H1-H2 indicates a more breathy phonation. The less abrupt glottal closing is correlated with the measure of H1-A3. A greater H1-A3 suggests a more breathy phonation. Although the spectral tilt could be a potentially vulnerable measure in that the amplitude of the second harmonic is affected by the first formant of the following vowel especially in a high vowel context, the spectral tilt measure has its own advantage as a convenient tool over the technique such as inverse filtering or fiberscopic observation of the vocal folds (Hanson, 1997). To minimize the effect of the filter (i.e., vowel type) in capturing the source characteristics of spectral tilt, correction methods were adapted that takes the bandwidth and the first formant frequency into consideration (Iseli, Shue, and Alwan, 2007).

The acoustic differences in the spectral tilt are useful in distinguishing a breathy vowel from a modal vowel which are phonemically contrastive in various languages. Huffman (1987) investigated the acoustic measures of glottal flow for vowels in Hmong which has both breathy and modal categories. The spectral analysis of Hmong vowels showed that the difference between the amplitude of H1 and the amplitude of H2 was consistently higher in the breathy vowel than in the modal vowel, suggesting that spectral tilt is a reliable way to show the breathiness contrast in Hmong vowels.

Wayland and Jongman (2003) used the measure of H1-H2 to demonstrate that the breathiness in Chanthaburi Khmer vowels was preserved only in female speech by being differentiated from the modal phonation in male speaker's vowels. H1-H2 was a successful measure in showing the distinction among modal, breathy and creaky vowel qualities in female speakers's productions in Santa Ana del Valle Zapotec (Esposito, 2004). In male speakers's three contrastive vowels, it was H1-A3 that

best capture the distinction between breathy, modal and creaky vowels, indicating that the phonetic output of phonation is manipulated by using different strategies of controlling glottal configurations, which was gender-specifically defined in the Santa Ana del Valle Zapotec speech community. Brunelle (2006) found that low register vowels in Eastern Cham were realized differently from high register vowels in terms of the breathy voice quality, by showing that H1-H2, H1-A1 and H1-A3 were higher in the low register.

Spectral tilt also played a role as a cue to the breathiness perception of English vowels. English listeners' rating of the vowel breathiness was highly correlated with an enhanced amplitude of the first harmonic in Hillenbrand, Cleveland, and Ericson (1994). The measure of H1-H2 worked best as an acoustic correlate of English listener's breathiness perception out of possible ways to quantify the spectral tilts.

1.4 Goals of research

This dissertation discusses the three questions introduced in Section 1.2 by examining children's stop production accuracy across four languages: Korean, English, Japanese and Greek. With the prosodic position confined to the word-initial stops, the selected target languages supply a useful set of stop laryngeal contrastive categories to explore three proposed questions. First, Korean stops with a three-way phonation-type contrast (i.e., lax, tense and aspirated stops) allow us to consider the question stated in Section 1.2.1 regarding how the mastery order of the stop categories is predicted based on their acoustic characteristics, when they do not belong to the three most common phonation-type categories (i.e., voiced, voiceless unaspirated and voiceless aspirated stops). Second, we investigate the mastery pattern of stop categories in Japanese where the voiceless stop is realized as these intermediate lag VOT values. This instantiates the question raised at Section 1.2.2. The relative order of mastering

Japanese voiced vs. voiceless stops will be discussed in a comparison with the order of mastering English voiced vs. voiceless stops which do not have this intermediate VOT values. Finally, Greek voiced stops are examined to explore the third question addressed in Section 1.2.3 that asks how the mastery of voiced stop would be affected by the their various language-specific acoustic details which are captured by the spectral characteristics of the prevoicing lead.

The investigation of these three questions begins by projecting the age of mastery based on the production accuracy of target stop categories in the database of children's production across languages. Chapter 2 describes this production accuracy of target stops in each language and relates the relative order of mastery with their VOT distributions set by adult productions of the language. This will tell us whether the VOT characteristics of the stop phonation-type categories adequately predict the production accuracy of children's stops or whether they give way to the puzzling patterns of children's production accuracy. Each of these three questions instantiated within the database is further discussed in the remaining chapters (Chapter 4-Chapter 6), taking into account the language-specific acoustic nature unique to the question. The unified theme of each chapter is to provide the language-specific evidence of how the mastery of stop laryngeal contrast would interact with the detailed acoustic characteristics of the category, aiming to provide more nuanced generalization of developmental universals in mastering stop consonants with a phonation-type contrast.

CHAPTER 2

THE TARGET LANGUAGES, THE PRODUCTION DATABASES AND ADULT VOT NORMS

This chapter describes the two databases of child and adult productions that were used in this dissertation and uses the databases to compare accuracy and VOT norms across five target languages. This crosslinguistic comparison of children's production accuracy patterns and of the VOT characteristics of adult productions of the categories lets us state the three questions introduced in Section 1.2 in a more concrete way.

The five languages from the database that are examined here are Cantonese, English, Greek, Japanese and Korean. This particular set of languages was chosen because it provides a sufficient variety of stop phonation-type contrast in word-initial position to make meaningful comparisons. Table 2.1 is a summary of the standard descriptions of the phonation-type contrasts in the five target languages. Specifically, Cantonese and English provide information about the phonetic characteristics of word-initial stops in languages with an aspiration contrast. Although the voiced stops of English differ from the unaspirated stops of Cantonese in optionally having voicing lead in word-initial position, English voiceless stops are characterized as obligatorily aspirated, closely resembling the aspirated stops of Cantonese. Japanese and Greek provide information about the phonetic characteristics of word-initial stops in

language	lead	short lag	long lag
Cantonese		/t, k/	/t^h, k^h/
English	/d, g/	[d,g] ∼ [t,k]	/t,k/ ([t^h, k^h])
Japanese	/d, g/	/t, k/	
Greek	/d, g/	/t, k/	
Korean		/t', k'/	/t, k, t^h, k^h/

Table 2.1: A summary of the standard descriptions of the phonation-type contrasts in the word-initial stop consonants of the five languages from the production database.

languages with a voicing contrast. Voiced stops in Japanese and Greek are characterized as truly voiced in word-initial position, realized as lead VOT values, while voiceless stops contrast with the voiced stops by having short lag VOT values. Korean word-initial stops provide information about the phonetic characteristics in a language that has other stop categories than the three used in 'true' voicing and aspiration contrasts.

2.1 Two databases

Two sets of cross-sectional data were used to make a crosslinguistic comparison of production accuracy across ages. One data set (Database I) provided the stop productions by Cantonese-, English-, Japanese- and Greek-speaking children. The other data set (Database II), which is larger in the terms of number of children, provided the productions by English-, Japanese- and Korean-speaking adults and children. Both databases are subsets of a larger production database collected in 2004 and 2006 in an NIDCD-funded Cross-Language Investigation of Phonological Development (`http://www.ling.ohio-state.edu/~edwards`). The Korean portion of the

data was gathered by the author in 2007 and added to the later database (sponsored by Targeted Investment fellowship from the Linguistics Department at Ohio State University awarded to the author). The two sets of cross-sectional data share the same target elicitation method, while they differ in the age range of child subjects, the number of consonant types elicited and the lists of target languages for the crosslinguistic comparison.

2.1.1 Materials

Target consonants were word-initial lingual stops in five following vowel contexts. For Greek and Japanese, the five contexts were the five vowel phonemes of the language – i.e., /i, e, a, o, u/ for Greek and /i, e, a, o/ and [ɯ] = "/u/" for Japanese. For Cantonese and English, the five vowel contexts grouped together phonemes with roughly comparable coarticulatory and phonotactic interactions with the preceding target consonant. For example, the short /i/ and long /iː/ were grouped together in the "/i/" context for Cantonese and the lax /ɪ/ and tense /iː/ were grouped together into the "/iː/" context for English (see Edwards and Beckman (2008) for the full list of groups). Three words were selected to elicit the target in each context. For example, Japanese /d/ vs. /t/ in the /o/ context were elicited in /doːnatsɯ/ 'donut', /do ŋguri/ 'rice bowl' and /doa/ 'door' vs. /toːfɯ/ 'tofu', /tomato/ 'tomato' and /tora/ 'tiger'. In English, word-initial /d/ vs. /t/ in /o/ context were elicited in /do.nət/ 'donut', /domz/ 'domes', /dor/'door' vs. /torn/ 'torn', /toad/ 'toad' and /tost/ 'toast'. The presentation order of the words was determined by a "randomizing" algorithm that insured that the words with the same sequence of target consonant and vowel are not neighboring with other in the list and that at most 2 words for any CV type were presented in either or two blocks. We ensured that the

target words were familiar to young children (see appendix A for the lists of words used to elicit the target lingual stops in Database I and Database II).

Selected target words were recorded by an adult female native speaker of each language to be presented to adult and child subjects to repeat after. The speaker pronounced each word five time using a child directed speaking style. Recorded tokens were carefully screened before being used in the production experiment in order to prevent children's mispronunciations caused by unintelligibly articulated stimuli. The screening was done by pretesting the production accuracy of five adult speakers' repetition after the prerecorded tokens. Tokens that at least four adult listeners repeated correctly (i.e., 80% accuracy) were selected as stimuli in the experiment If not enough of the original five tokens of a target word met the criterion, the speaker rerecorded more tokens that were again screened in a repetition task. As a byproduct of this screening procedure, stop productions by three Cantonese- and three Greek-speaking adults in Database I could be acoustically analyzed to show the crosslinguistic VOT distributions at Figure 2.3.

In addition to the audio stimuli, culture-appropriate pictures of target words were prepared to be presented together with the audio stimuli in the production experiment. Two of the pictures used in the task are shown in Figure 2.1.

2.1.2 Subjects

Database I collected target stop productions by children speaking Cantonese or English aged 2;0-3;11 and Greek and Japanese aged 2;6-3;11. Database II covers a wider age range (2;0-5;11) of child subjects speaking English, Japanese and Korean. This larger database also has the stop productions by 20 adult speakers aged 18-30. The age distributions of the child subjects for each of the two databases are given in Table 2.2 and Table 2.3.

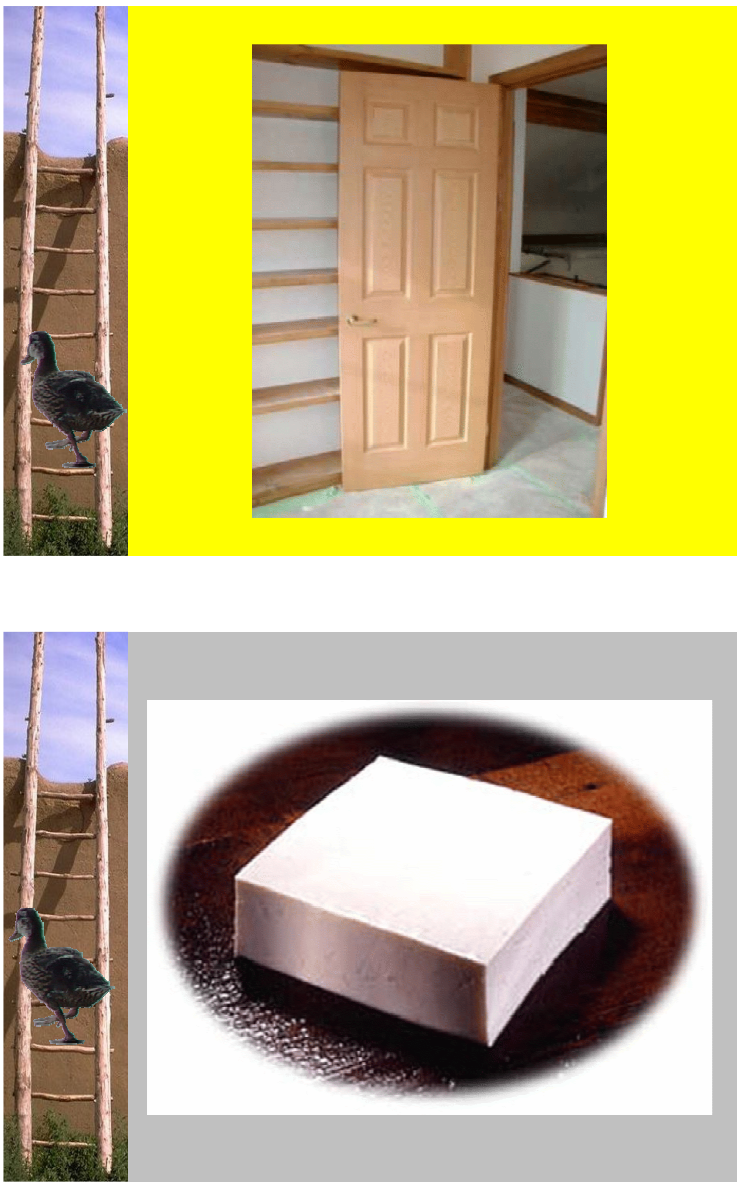

Figure 2.1: Pictures used to prompt the target words: Japanese /doaː/ ‘door’ and /toːfɯ/ ‘tofu’. The duck at the left side climbs the ladder as the child repeats after the target word, which is devised to motivate child subjects to finish the task.

	English		Japanese		Greek	
age group	girls	boys	girls	boys	girls	boys
2;0-3;0	8	3	9	3	5	4
3;0-4;0	5	7	5	3	5	4

Table 2.2: The age distributions of child subjects speaking English, Japanese and Greek in Database I.

	English		Japanese		Korean	
age group	girls	boys	girls	boys	girls	boys
2;0-3;0	6	7	6	7	12	9
3;0-4;0	7	4	7	5	11	9
4;0-5;0	8	6	6	8	7	8
5;0-6;0	6	6	5	7	2	6

Table 2.3: The age distributions of child subjects speaking English, Japanese and Korean in Database II.

Both adults and children were recruited and recorded in their native lands (the Cantonese speakers in Hong Kong, the English speakers in Columbus, OH, USA, the Japanese speakers in Tokyo and Hamamatsu Japan, the Korean speakers in Seoul, Korea, and the Greek speakers in Thesaloniki and Athens, Greece). All the child and adult subjects passed a hearing screening.

2.1.3 Task

Target consonants were elicited using a picture-prompted word-repetition task. On each trial, a computer program presented the picture on the monitor and played the recorded audio stimulus naming the picture through an external speaker. Subjects were asked to repeat after the audio presentation. Their repetitions were recorded

using a CD recorder and a head-mounted microphone(Shure SM10A) for the collection of Database I and a flash card recorder (Marantz PMD660), and a hand-held microphone(AKG C5900m) for the collection of Database II. Experimenters were native speakers of the target language. They were instructed to elicit the words if a child balked at repeating after the computer voice and to not use a live-voice prompt. When children made multiple repetitions during the sessions, we include only the first transcribable tokens for the analysis. The whole task consisted of four sessions, two real-word repetition sessions followed by the two nonword repetition sessions (I used children's real-word repetitions only). The total duration of word-repetition task session differed from child to child, but did not exceed 20 minutes overall.

Child subjects were also given an articulation test (Goldman-Fristoe articulation test for English children (Goldman and Fristoe, 2000) and comparable tests for Japanese (Nihon Chōin Gengo Hakasekai and Nihon Onsei Gengo Igakukai, 1994) and Korean children (Kim, 1996)), and a receptive vocabulary test in order to make sure their language development is on track. Children were rewarded with a small toy after each session, and adults were paid for their participation. Recordings were made in child care centers or in the children's homes for the youngest children.

2.2 Analysis

2.2.1 Transcription

The accuracy of phonation-type contrast was estimated based on a single native speaker transcriber's transcription. Transcribers of each language were trained phoneticians.

The transcription of each target production was done in two steps. As a first step, the transcriber decided whether the consonant production was on target or not

by coding it as '1' for correct, 'V' for correct except for voicing or '0' for incorrect. As a next step, the transcriber provided an alphabetic transcription of those productions that were judged as incorrect. WorldBet symbols were used for the transcription (Hieronymus, 1994). The transcribers for the English and Japanese productions in the two databases were different.

The evaluation of accurate phonation-type relied on those substitution transcriptions of children's stop productions. While non-plosive productions were regarded as errors, place of articulation errors were ignored for the current analysis, as long as the phonation-type categories were correct. For example, [g] or [dʃ] for a target /d/ in English was counted as correct for voicing despite the error in the place of articulation. We excluded the tokens of deletion, distortion and non-plosive productions from the analysis.

2.2.2 Age of mastery

The age of mastery of each stop phonation-type category was assessed by predicting the age of reaching 75% accuracy in an apparent time analysis using a mixed effects logistic regression model. The model predicts the transcribed accuracy of each stop as a function of children's age, positing the random effect of speaker. Equation 2.1[1] shows the formula of this model.

$$(2.1)\quad log(\frac{correct}{1-(correct)}) = \beta_0 + \beta_1\textbf{Age} + \beta_2\textbf{Consonant Type} + \gamma\textbf{Speaker}$$

A separate model was built for each language. The age for which the model predicts a 75% accuracy rate is defined as the age of mastering the target stop phonation

[1] R.code:`lmer(accuracy~age+targetC+(targetC|subj),data=accuracyData,family=binomial)`

type. These projected ages of mastery allow us to make a crosslinguistic comparison of relative order of mastering stop phonation-types.

The criterion of voicing mastery was set as 75% on the basis of criteria chosen in various norming studies (Smit, Hand, Freilinger, Renthal, and Bird, 1990; So and Dodd, 1995; Amayreh and Dyson, 1998) where the age of 'acquisition' or 'mastery' is defined by the percentage of children who produce the target speech segment correctly. Various studies have adopted different percentages to indicate the age of mastering speech segments. For example, Templin (1957) defined the age of 'acquisition' as the age at which 75% of children in a particular age group produce the target segment correctly, while So and Dodd (1995) adopted a more rigid criterion of 90% of children making a correct production of the target segment. Sander (1972) defined several criteria that differentiate consonant accuracy in different prosodic positions. Thus she defines (1) the 'age of customary production' as the age when 50% of children produce the consonant correctly in at least two prosodic positions, (2) the 'age of acquisition' as having 75% of children produce the consonant correctly in all positions tested, and (3) the 'age of mastery' as the age when at least 90% of children produce the consonant correctly in all positions tested.

Note that each of these prior norming studies used a picture-naming task which elicited the consonants in only one or two words in the different target word positions. The criteria, therefore, have to be group criteria, which count the number of children when they count the number of productions. Since we have as many as 15 tokens of each consonant per child, we can define the criterion of 75% correct productions child by child to get a fine-grained estimate in terms of months of age rather than in terms of coarser age groups.

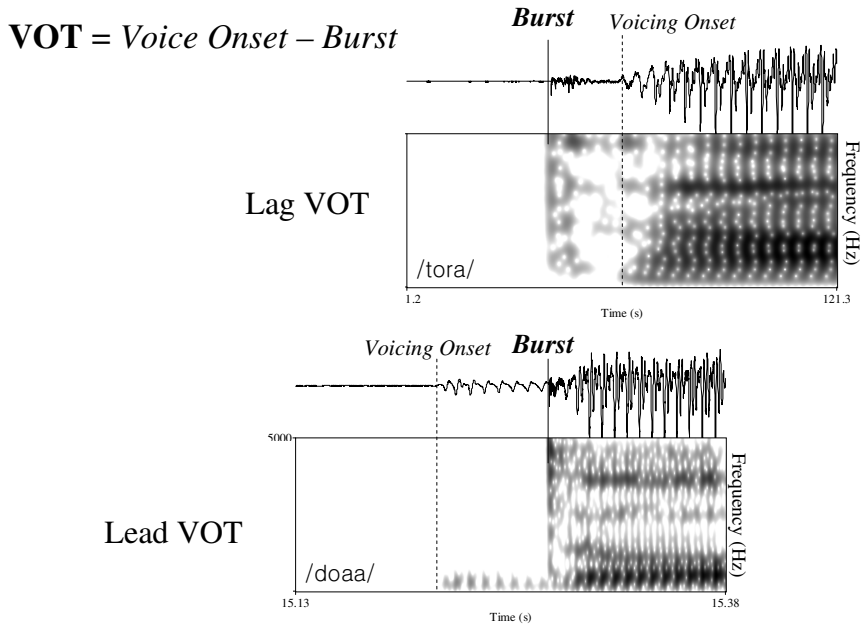

Figure 2.2: Illustration of VOT measurement. The top figure shows an example of lag VOT (Japanese /tora/ 'tiger') and the bottom figure shows an example of lead VOT (Japanese /doaː/ 'door').

2.2.3 Voice Onset Time

VOT was measured by subtracting the time of burst (i.e., the beginning of an abrupt energy rise after the closure), from the time of the first indication of voicing, evident from the voicing bar in the spectrogram, as well as a visible initiation of the first regular cycle of periodicity in the waveform. Figure 2.2 shows these landmarks for a production with lag VOT and for a production with lead VOT. The measured VOTs in milliseconds were converted to log scale.

2.3 Results

2.3.1 Voice Onset Time

Figure 2.3 shows the distributions of VOT values measured in adult productions of word-initial stops across the five target languages. The top four panels show the VOT ranges for the two categories of Cantonese, English, Japanese and Greek, and the last panel shows the VOT ranges for the three categories of Korean. The histograms of VOT distribution are based on three Cantonese- and three Greek-speaking adults' stop productions in Database I, and fifteen English-, twenty Japanese- and twenty Korean-speaking adults' productions in Database II. Because prior studies have established place-related VOT differences, the VOT distributions of coronal stops are plotted separately from those of dorsal stops. The longer VOT in dorsals might be due to the slower release of a dorsal closure, which could delay the venting of air pressure to make the transglottal pressure difference necessary for the initiation of vocal fold vibration (Cho and Ladefoged, 1999; Maddieson, 1996).

The figure shows five different VOT distributions across the five target languages. Two of the languages showed fairly well separated values exemplifying the three canonical types. First, the aspirated category in Cantonese was consistently realized with a long VOT lag, whereas the unaspirated category was realized with a short VOT lag. Thus, the unaspirated and aspirated stops in Cantonese were fairly well separated from each other along the VOT continuum, a result which is consistent with the findings in Lisker and Abramson (1964).

Second, the English stop VOT distributions of voiced versus voiceless stops resembled those of the unaspirated versus aspirated stops in Cantonese, in that VOT values of word-initial voiceless stops in English are in the long lag range, like the aspirated stops in Cantonese. The voiced stops in English were realized predominantly

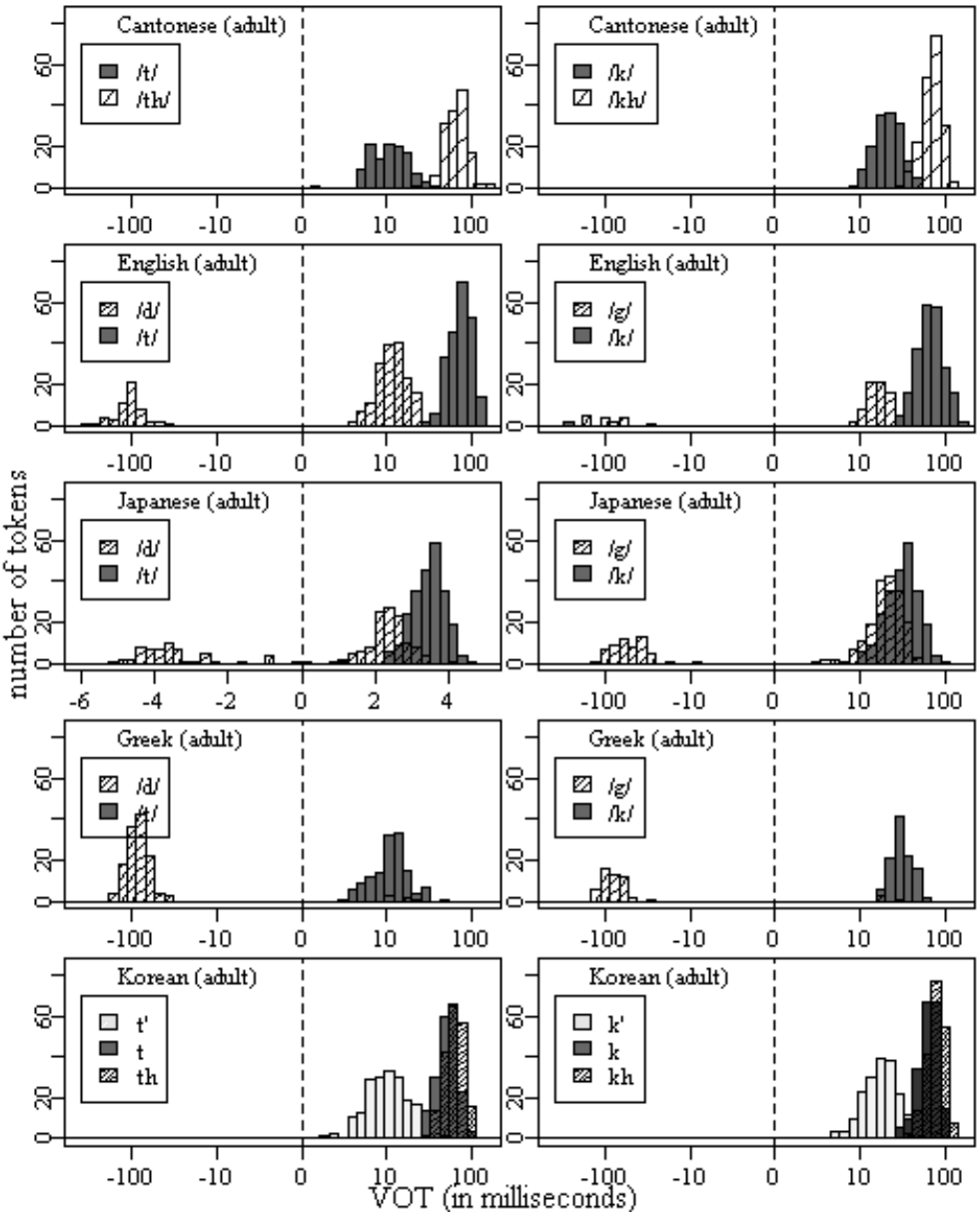

Figure 2.3: The VOT distributions of adult productions of word-initial stops in Cantonese, English, Japanese, Greek and Korean.

with short lag VOT values, with only a small number of tokens with lead VOT. That is, only 23% (73/314) were 'true' voiced stops.

Third, the Japanese stops show a voicing contrast that patterned differently from the English stops, such that VOT values for the Japanese voiced stop values overlap with those of the voiceless stops. This overlap appears to come from two characteristics of Japanese stops. On the one hand, Japanese voiced stops were realized with short lag VOT as well as with lead VOT. 79.92% (309/418) of voiced stops had short lag VOT values. The pattern is consistent with the findings in Takada (2004b). On the other hand, the Japanese voiceless category has a VOT distribution peak that is intermediate between the peaks of the short lag VOT for English voiced stops (and Cantonese unaspirated stops) and the long lag VOT for English voiceless stops (and Cantonese aspirated stops). This kind of intermediate lag VOT category has been documented in the study of Japanese language (Riney et al., 2007) and other languages such as Hebrew (Raphael et al., 1995) and Canadian French (Caramazza et al., 1973).

Forth, the voiced category in Greek was separated from the voiceless stop by having typical lead VOT values with only a few voiced tokens in the short lag range, which is unlike the English and Japanese VOT patterns.

Finally, the Korean stops show an even more radical overlap, because the lax stops, like the aspirated stops, are a long lag category. Korean has been described as showing at least some VOT overlap since earlier studies (Lisker and Abramson, 1964; Kim, 1965; Han and Weitzman, 1970). However, these document an overlap only between the lax and tense stops, in the short lag VOT range, with only the aspirated stops showing values in the long lag VOT range and clearly separated from the other two phonation types. The current data show a very different pattern of VOT overlap, where by the tense stops in the short lag VOT range are clearly separated from lax

and aspirated stops at the long lag VOT range. This different pattern of VOT overlap in Korean stops has been described as a diachronic sound change in Silva (2006) and Kang and Guion (2008). Whereas older Korean speakers have VOT distributions that make the lax stops an intermediate VOT type which overlaps more with the tense stops, Korean speakers born after the 1970s tend to produce lax stops with long lag VOT values that separate them from tense stops but merge them on the VOT dimension with aspirated stops.

The overlapping VOT range in Korean and Japanese means that VOT cannot serve as the sole acoustic dimension distinguishing the stop phonation-type categories in these two languages. This implies that there must be other relevant acoustic dimensions that disambiguate the stop phonation-type categories with the overlapping VOT values in these languages. Section 1.3 listed some of the dimensions that might be relevant and Chapters 3-5 will explore them in some detail in the adult productions shown in Figure 2.3

2.3.2 Transcribed accuracy

Figure 2.4 and Figure 2.5 present the patterns of transcribed accuracy across the target languages. Recall that we analyzed the children's productions that were first attempts to repeat the target words and that were transcribed as plosives – i.e., excluding deletions, distortions and substitutions of sounds other than non-plosives. In each figure, the percentage of each individual child's target stops that met these criteria that were judged to have the target phonation type is plotted as a function of the child's age in months. The regression curves overlaid on the data points are generated by a mixed effects logistic regression model as shown in Equation 2.1. The dotted horizontal lines indicate a 75% accuracy rate, which was our reference rate for defining mastery of the category. There are three noteworthy patterns.

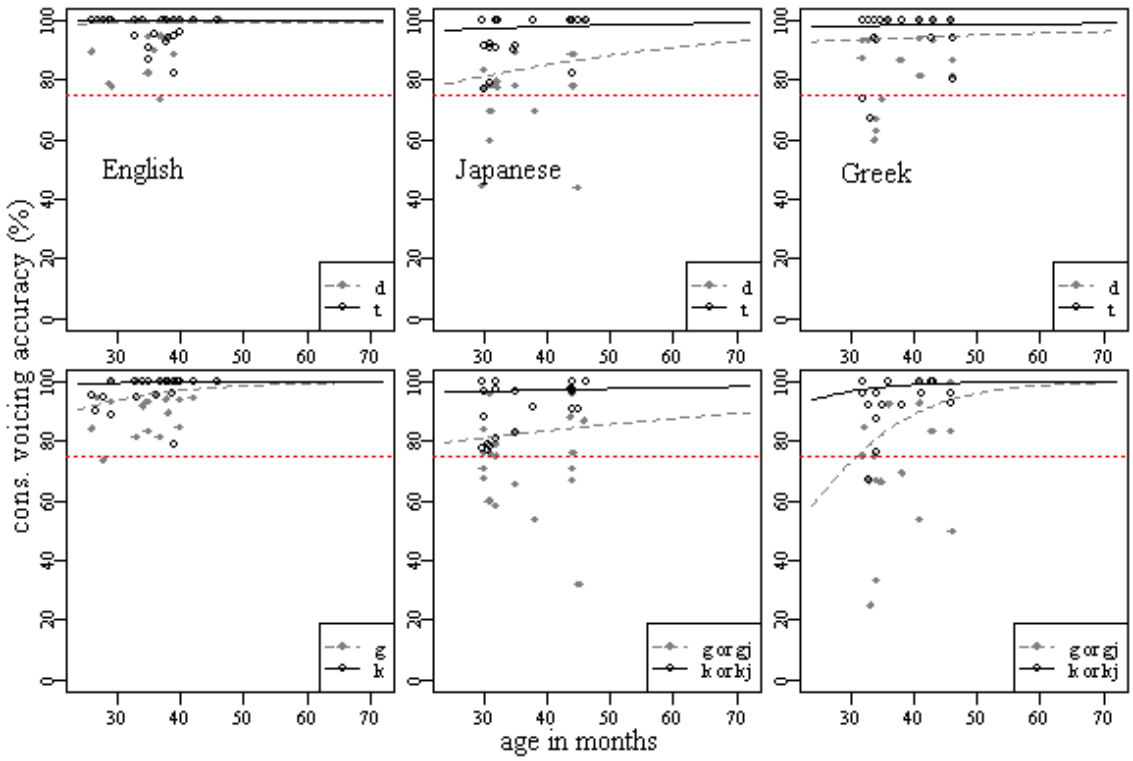

Figure 2.4: Transcribed accuracy of phonation-type categories for children's word-initial stops in English, Japanese and Greek in Database I.

First, There was a difference in accuracy between Japanese voiced and voiceless stops well past the age when English-speaking children have mastered both types in their language. Consistent with prior findings from longitudinal studies (Macken and Barton, 1980a; Zlatin and Koeigsknecht, 1975) and from cross-sectional norming studies (e.g., Smit et al., 1990), both voiced and voiceless stops in English were produced by children with more than 75% voicing accuracy before 24 months. Japanese voiceless stops also have more than 75% voicing accuracy even before 24 months. By contrast, Japanese voiced stops had a lower voicing accuracy than voiceless stops in the same age range as indicated by the mixed effects models for both databases.

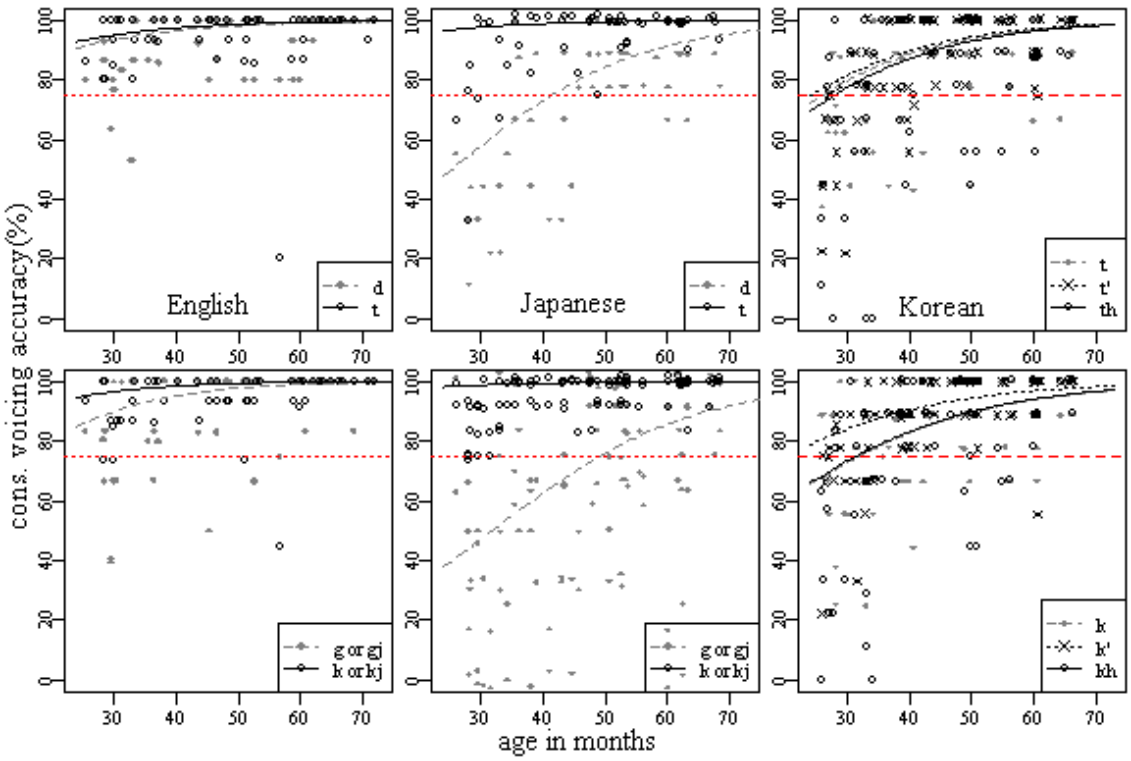

Figure 2.5: Transcribed accuracy of phonation-type categories for children's word-initial stops in English, Japanese and Korean in Database II.

Moreover, although the regression lines for the voiced stops are above the 75% reference line for both the coronals and the dorsals in Database I, the curves for the mixed effects models for Database II cross the 75% voicing accuracy lines for coronal voiced stops at about 43 months, and for dorsal voiced stops at 50 months. This lower accuracy for voiced stops conforms to earlier crosslinguistic findings of later mastery of 'true' voiced stops in Spanish (Macken and Barton, 1980b), French (Allen, 1985) and Thai (Gandour et al., 1986).

Second, Greek voiced stops in Database I suggest a pattern of earlier mastery of voiced stops as compared to Japanese voiced stops. The logistic regression curve for the voicing accuracy of Greek voiced coronal stops was well above 75% before 24

months, and the curve for the Greek voiced dorsals converges forward the curve for voiceless dorsals by 46 months. This early mastery of Greek voiced stops estimated by the consonant production accuracy is unlike Japanese stop voicing mastery in this database and deviates from the crosslinguistic trend of late mastery of voiced stops.

Third, Korean tense, lax and aspirated stops were all mastered at a relatively early age albeit not as early as the English voiced (unaspirated) and voiceless (aspirated) stops. Based on the regression curves from the mixed effects models, the ages of 75% voicing accuracy are projected to be 24 months, 26 months and 29 months for /t'/, /t/ and /t^h/, respectively and 20 months, 33 months, and 33 months for /k'/, /k/ and /k^h/, respectively. While the estimated ages of mastery based on the transcribed accuracy were not greatly different between tense stops and the other two categories, the accuracy of Korean tense stops appears to reach the 75% criterion slightly earlier than lax and aspirated stops. This is consistent with findings in some other smaller studies of children who are younger than those recorded here.

Moreover, the substitution patterns in the Korean children's error productions in Database II also support the idea that the tense stop appears before the aspirated stop, because a tense stop is commonly substituted for a target aspirated stop in the younger children's productions. Figure 2.6 shows the percentage of errors child-by-child where a lax or an aspirated target stop was produced as a tense stop (tense substitutions/total errors) as a function of the child's age in months. The two curves overlaid on the scatter plots show the results of a mixed effects logistic regression model where the tense substitution category of error productions was regressed against the child's age and the target phonation type as fixed effects (the individual child was modeled as a random effect in this mixed effects model). The equation is shown in Equation 2.2 (note that the data input set for this analysis is only the

error productions for lax or aspirated targets). The size of the plotting character indicates the number of relevant errors that the child made in the production accuracy analysis. According to the model, there is greater likelihood for a tense type to be substituted for a target aspirated stop in younger ages and the use of the tense type decreased over time. While the curve for the target lax stop shows that children were more likely to substitute an aspirated stop in an errored production of a target lax stop, overall there are fewer such substitutions (i.e., more tense substitutions for the youngest children).

$$(2.2)\quad log(\frac{tense}{1-(tense)}) = \beta_0 + \beta_1 \mathbf{Age} + \beta_2 \mathbf{Consonant\ Type} + \gamma \mathbf{Speaker}$$

2.4 Three puzzles

The three noteworthy patterns in the transcribed accuracy rates examined across the target languages in Database I and Database II were the original observations that prompted the questions raised in Section 1.2. Summarizing those questions again, I want to know the extent to which the VOT characteristics of stops in a language predict the mastery patterns for stop laryngeal contrasts in first language acquisition. In this section, we elaborate on the three questions outlined in Chapter 1 by reviewing the prior literature on Korean, Japanese and Greek, and relating what is known about the articulation of the contrasts in question to the results reported in Section 2.3.

2.4.1 Korean

The accuracy patterns for Korean stops in the database let us address the questions of when and how children master the stop phonation-type categories when a language

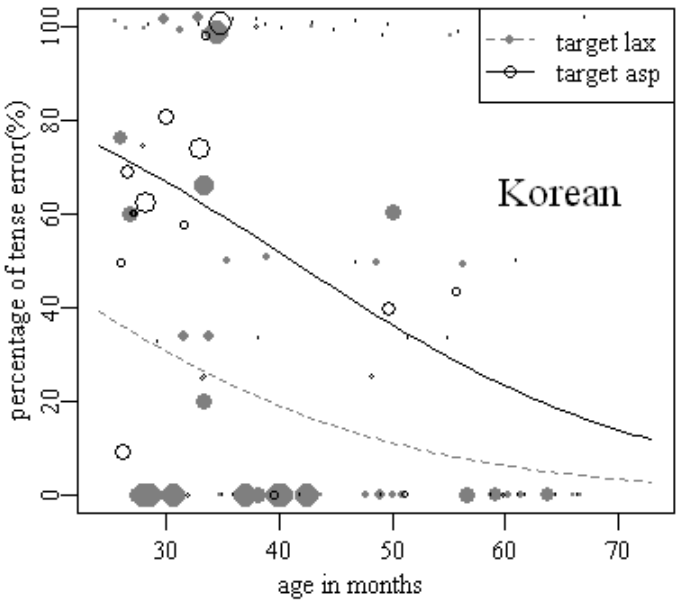

Figure 2.6: Scatterplot of percentages of substituted tense tokens out of errored productions as a function of the child's age in months. The curves are the probability of being tense estimated by the mixed effects logistic regression model that formulates the relationship between the percentage of substituted tense tokens and the child's age.

has a contrast which is not defined as either a voicing or an aspiration contrast. Two contradictory predictions can be made with respect to the mastery order among lax, tense and aspirated stops based on the phonetic nature of the stop categories in Korean. On the one hand, we might predict that the tense stop would be mastered later than the other categories because it is a rare phonation type across languages, which is made with a complex sequence of glottal gestures combined with a vocal tract condition of tenseness, whereas the lax stop would be mastered earlier due to the lack of positive laryngeal activities in the production (Kagaya, 1974; Hardcastle, 1973; Hirose et al., 1974; Dart, 1987).

On the other hand, we might predict the early mastery of Korean tense stop for the same reason that the unaspirated voiceless stop is predicted to appear first in children's productions across languages. Figure 2.3 shows that in adult productions the tense stop occupies the short lag range and has little overlap with the VOT values for lax and aspirated stops. If Korean children's early stop productions have short lag VOT, as predicted from young children's stop productions across languages, the tense stop will be the adult category that corresponds most closely to the short lag stops produced by children. Unless other acoustic parameters besides VOT are more important in differentiating tense stops from lax and aspirated stops, children's early stop productions are likely to be assimilated to the adult tense stop.

The second prediction accords more with accuracy rates in the current database. According to the transcription analysis, tense, lax and aspirated stops were all mastered before three years, although the tense category seems to be mastered somewhat earlier than the aspirated category. As noted earlier, this result is in keeping with the previously described order of mastery among stop phonation-type categories in cross-sectional and longitudinal studies of phonological development in Korean stops. For example, Kim and Pae (2005) conducted norming studies for consonants at various prosodic conditions with Korean-speaking children aged 2;6 to 6;5. They found evidence that tense stops were mastered before lax or aspirated stops, in that word initial /p'/ and /t'/ were produced correctly by 95% of children before 2;6. All three were mastered by 75% of children in the age group from 3;1 -3;6. Kim (2008) transcribed multiple repetitions of target word-initial stops per child. Tense stops /p', t', k'/ were produced with 75% accuracy by more than 75% of children at 2;6, while the same accuracy for aspirated stops /p^h/, /t^h/ and /k^h/ and lax /t/ were not acheived until 3;0 and 4;0. In a longitudinal case study by Jun (2007), word-initial tense stops were recorded to appear at 17 months, before aspirated (18 mo.) and lax (20 mo.).

In order to resolve this puzzle of mastering three types of Korean stops, the Korean children's stop productions are compared to the adult productions in terms of various acoustic characteristics including *f0*, H1-H2 and VOT. The acoustic characteristics of Korean children's stop productions are also related to the categories transcribed by the native transcriber to assess the role of different acoustic cues in determining the tense, lax and aspirated categories in Korean.

2.4.2 Japanese

The accuracy rates for the Japanese children's stops in Figure 2.4 and Figure 2.5 follow the crosslinguistically common pattern of mastering voiceless stops before voiced stops. However, while this order is consistent with what has been documented in languages such as French, Spanish and Thai, where there is a true voicing contrast, it is puzzling when the VOT characteristics of Japanese voiced stops are taken into consideration.

The adult VOT distributions in Figure 2.3 show that voiced stops in Japanese are realized not only with lead VOT values but also with short lag values, which then overlap with the intermediate VOT values for voiceless stops. Given that the late mastery of voiced stops across languages is attributed to the motoric demand of producing lead VOT, the late mastery of Japanese voiced stops is puzzling. Japanese-speaking children can realize voiced stops with short lag VOT values, which would not be subject to these motoric constraints. Thus, the late mastery of Japanese voiced stops is not predicted by the VOT characteristics of Japanese stops, leading to the suspicion that there might be acoustic parameters other than VOT that are important in differentiating voiced stops in Japanese. To resolve this puzzling accuracy pattern in Japanese voiced vs. voiceless stops, therefore, we need to examine which acoustic parameters contribute to the characterization of the contrast in Japanese.

In Chapter 4, I investigate the acoustic characteristics of the Japanese voiced stops along multiple acoustic dimensions, using measures of breathiness and *f0* as well as VOT. We hypothesize that the voicing judgments reported in this chapter are more effectively described when multiple facets of the categories are taken into consideration. Specifically, the productions by Japanese- and English- speaking adults and children in Database II are analyzed and the relationships between multiple acoustic properties and the voicing judgment are explored using mixed effects logistic regression models. Chapter 5 then reports results of a perception experiment following up on the analysis of the production database. By comparing the Japanese and English models, we can begin to understand how the existence of the intermediate VOT type affects the interpretation of the voicing contrast in languages where the voiced stop can be realized with short lag VOT values as well as with lead VOT.

2.4.3 Greek

The third puzzling pattern in the database is the very high accuracy rate for the Greek voiced stop. The regression curves in Figure 2.4 suggest that the voiced stops of Greek are mastered even before 3 years, which is exceptional given the crosslinguistic trend of very late mastery for the "true" voiced stop.

The adult VOT distributions make this high accuracy puzzling because Greek voiced stops have typically long lead VOT, and do not seem to allow the short lag VOT variant depicted in Figure 2.3.

One possible explanation of this puzzle invokes language-specific characteristics of the voicing lead which are related to the fact that Greek voiced stops developed historically from clusters of nasals followed by voiceless stops. Arvaniti and Joseph (1999) describe the Greek voiced stop as still having more or less prenasalized variants, which are conditioned by speech style, speaker gender and prosodic position. Given

this socio-historical background, it is possible that Greek children will be perceived as accurate even when they produce their voiced stops with a fairly high degree of nasalization in the voicing lead interval. The nasalized variants should make it easier for children to control the supraglottal pressure by using the naso-pharyngeal port as an additional air channel. We hypothesize that children maintain a lower supraglottal pressure relative to subglottal pressure by leaking through the naso-pharyngeal port.

In order to test this hypothesis, we examine spectral characteristics of the voicing lead interval using measures adapted from descriptions of contrastively pre-nasalized stops found in Fijian (Maddieson, 1989) and Moru (Burton et al., 1992). The degree of nasality in Greek voiced stops will be assessed by comparing nasal murmurs in the nasal consonants of the language with the voicing lead interval in voiced stops. The Greek patterns are also compared to the pattern for the same measures for voiced stops and nasal consonants in Japanese, another language in the database with lead VOT voiced stops.

CHAPTER 3

THE THREE-WAY CONTRAST IN KOREAN

This chapter explores the relationship between the order in which children master the three different stop phonation types of Korean and the role of various acoustic characteristics in cuing the contrast. As stated in Lisker and Abramson (1964), the three-way phonation-type contrast in Korean stops (i.e., lax versus tense versus aspirated stop) is not successfully described just by VOT due to the overlapping range among categories. The Korean pattern contrasts with Thai, where the effective differentiation along the VOT continuum among three largely non-overlapping distributions for voiced versus voiceless unaspirated versus aspirated stops leads to a natural explanation of why Thai-learning children master /t/ first, then /t^h/, and /d/ last. Since VOT does not neatly differentiate Korean /t'/ from /t/ from /t^h/, we need to ask what other cues are available for the differentiation among the three Korean stop types and what explanations these cues offer for the observed order in which the stops are mastered.

3.1 Prediction

3.1.1 Production accuracy and VOT

The data reviewed in Figure 2.5 showed that the accuracy of tense, lax and aspirated stops in Korean reached the 75% mastery criterion before age three. Moreover, the

substitution patterns in Figure 2.6 showed that younger children were transcribed as substituting a tense type more often than a lax or an aspirated type in their productions that were judged to be inaccurate. These observations are consistent with findings from prior normative studies of Korean where the consonant accuracy was evaluated by native speakers' transcription (Kim and Pae, 2005; Kim, 1996).

The early mastery of tense stops is puzzling when the phonetic characteristics of the Korean tense stop are considered. The Korean tense stop has short lag VOT. However, it is not at all like the 'default' or 'unspecified' voiceless unaspirated stop of Cantonese, Hindi or French. Rather, it is specified as having a tightly adducted glottis (the posture for creaky voice or glottal stop) that prevents vocal fold vibration during the closure (Hirose et al., 1974; Kagaya, 1974). There is also substantial vocal tract wall tension, which accounts for a higher intraoral pressure despite the lower air flow at the oral constriction release than in the lax stop (Dart, 1987). These complex glottal and supraglottal settings for Korean tense stops are associated with acoustic differences in various acoustic dimensions other than VOT. There is a higher fundamental frequency (*f0*) in the following vowel that differentiates tense and aspirated stops from lax stops, and the tenseness of the vocal cords also creates voice quality differences that distinguish the tense from the lax and aspirated stops (Kim, Beddor, and Horrocks, 2002; Cho, Jun, and Ladefoged, 2002). The early mastery of the tense stop is puzzling because we would expect the complex phonetic characteristics of tense stops to delay children's mastery of the category. The goal of this chapter is to investigate the role of the various acoustic parameters in predicting the transcribed production accuracy of the Korean children's tense, lax and aspirated stops.

3.1.2 Other acoustic parameters

3.1.2.1 Spectral tilt

The different glottal configurations of the three Korean stops affect the following vowel quality in such a way that the vocalic onset after the tense stop has a pressed quality, whereas the vocalic onset after the lax stop has a breathy quality and only the aspirated stop has an unspecified or modal voice quality. When H1-H2 (the amplitude difference between the first and second harmonics) is measured to capture the voice quality in the vowel just after lax, aspirated, and tense stops, the lax stop is characterized as having the greatest H1-H2 value and the tense stop as the smallest (i.e., negative) H1-H2 value. This acoustic parameter of H1-H2 can distinguish the tense from the lax and the aspirated stops in Korean in adult speakers' productions (Cho et al., 2002; Kim, 2008). In perception, this H1-H2 parameter seems to cue the stop phonation-types in that a higher H1-H2 after the stop consonant biases Korean listeners against the tense stops when they listen to cross-spliced CV stimuli with conflicting phonation-type cues (Kim et al., 2002).

Kang and Guion (2008) found that the H1-H2 parameter is used for enhancing the contrast between the tense stop and the other two stop types when the speech style changes from casual speech to clear speech. The H1-H2 values of tense stops are lower in clear speech than in casual speech, whereas the H1-H2 values of lax and aspirated stops do not differ between the two speech types.

According to Kim (2008), Korean-speaking children's stops also show these characteristic H1-H2 differences among tense, lax and aspirated stops, by having the smallest (i.e., negative) H1-H2 at the vocalic onset after tense stops even as early as 2 years 6 months. The mean H1-H2 values of tense stops produced by even the

youngest children (2;6) in this cross-sectional study are significantly lower than those of lax and aspirated stops.

3.1.2.2 Fundamental frequency (*f0*)

The three different phonation types of Korean stops are distinguished by *f0* in addition to VOT. Vowels after lax stops are differentiated from vowels after aspirated stops and tense stops by having a lower *f0* (Cho et al., 2002; Kang and Guion, 2008; Kim et al., 2002; Wright, 2007). The lower *f0* values after lax stops can resolve the ambiguity that arises from the VOT overlap between lax and aspirated stops in phrase-initial position. This *f0* difference in Korean stops is phonologized in the intonation structure of Seoul Korean such that the tense or aspirated phonation type triggers a high tone at the beginning of the accentual phrase, whereas the lax type triggers the low tone initial accentual phrase variant (Jun, 1993, 1998). In Seoul Korean speakers' stop productions, then, aspirated stops are observed to have the highest *f0* and lax stops the lowest *f0* just after release (Kang and Guion, 2008; Kim, 2008).

While different studies of Korean stop *f0* have consistently found that the lax stops have a lower *f0* than the tense and aspirated stops, the *f0* value of tense stops in relation to the *f0* of aspirated stops appears to vary among studies. This discrepancy in the findings might be due to the speaker's age. Younger speakers of Seoul Korean (born after the 1970s) produce tense stops with a lower *f0* than aspirated stops (Kang and Guion, 2008; Kim, 2008), whereas older speakers do not make a difference between the *f0* values of tense and aspirated stops (Cho et al., 2002). When Cho et al. (2002) tested speakers in their 50s, 60s and 70s, the *f0* values between tense and aspirated stops were not significantly different, although the lax stops were still distinguished from them by their lower *f0*. Age-related differences in the use of acoustic parameters are addressed by Kang and Guion (2008), who compared the

different enhancement strategies between younger and older speaker groups in terms of the acoustic parameters in encoding the stop phonation-types. The younger group of speakers, born after 1977, increased *f0* of aspirated stops in clear speech relative to conversational speech to enhance the phonation-type contrast, while older group of speakers increased VOT differences among the three stop categories. Kang and Guion (2008) interpret their results as reflecting the diachronic change to longer VOT values for the lax stop in Korean.

Several previous studies also suggest that children use the *f0* parameter in distinguishing the phonation-type contrast in Korean. For example, Kang (1998) found, in her case study with one Korean speaking child (2;8), that the vocalic onset *f0* values of tense and aspirated stops were higher than those of lax stops, and this difference, along with a short lag VOT, discriminated tense stops from aspirated and lax stops. Another longitudinal case study by Jun (2007) also reported that, from 18 months of age, the child made a low *f0* for lax stop relative to the *f0* for tense and aspirated stops, suggesting that the suprasegmental property (i.e., *f0*) is mastered before the segmental property (i.e., VOT). A cross-sectional study by Kim (2008) confirmed this early use of *f0* in Korean children's productions in distinguishing lax from tense and aspirated stops. Children aged 3;0-4;0 produced lower *f0* values for lax stops and higher *f0* values for aspirated stops with little or no overlap in *f0* values. The *f0* means of older children's stops were higher for the aspirated and tense stops in that order, and lower for the lax stops, reflecting the adult pattern of *f0* realization. The youngest group of children (2;6) showed *f0* values in tense stops that were not effectively separated from those in lax stops, and the *f0* values in aspirated stops also overlapped with the *f0* values in lax stops.

However, *f0* appears less effective in cuing the percept of the tense type. While high *f0* characterizes the tense stop in production, it may not function to distinguish

tense stops from the other two kinds of stops in perception. In a study by Kim (2004), Korean adult listeners' identification of the tense stop did not correlate with *f0*, whereas the identification of lax versus aspirated stops was strongly correlated with *f0*. These results suggest that VOT is a sufficient acoustic cue to distinguish the tense type from the other two types.

3.1.3 Hypothesis

If Korean adult listeners judge the tense type of stops based on VOT, with little influence from other acoustic parameters such as *f0* and H1-H2, a young children's short lag VOT tokens may be heard as tense stops, even though the child is not producing the voice quality that characterizes these stops in adult productions. Thus, we hypothesize that the tense stop is mastered before the lax and the aspirated stop because VOT is a sufficient acoustic parameter for cuing the tense category in Korean. In this chapter, we test this hypothesis by investigating which acoustic parameter in Korean stops best determines the accuracy of stop phonation-type in children's productions when the stop tokens are judged by a native transcriber. Specifically, the stop productions by Korean-speaking children and adults from Database II are analyzed in terms of the acoustic parameters of *f0*, H1-H2 and VOT and regression models are used to examine the relationship between the acoustic properties of each token and the transcribed phonation type (for children) or the target phonation type (for adults).

	Korean	
age group	girls	boys
2;0-2;11	12	9
3;0-3;11	11	9
4;0-4;11	7	8
5;0-5;11	2	6
adults	10	11

Table 3.1: The age distributions of child and adult subjects speaking Korean in Database II.

3.2 Method

3.2.1 Materials, subjects and tasks

The Korean subset of Database II was analyzed. The description of the materials, subjects, and task were presented in Chapter 2. Table 3.1 repeats the age distribution of child participants in Database II.

Out of all the children's tokens used in the accuracy analysis, the tokens transcribed as stops, in which the burst and the voicing onset were measurable, were included in the acoustic analysis. That is, this data set for the acoustic analysis consists not only of tokens judged as '1' or 'V' but also those tokens judged as '0' where the transcriber identified a stop substitution, such as lax [t] for a target /k/. Table 3.2 provides the number of tokens from the child and adult participants that were used in the acoustic analysis.

target consonant	2;0-2;11	3;0-3;11	4;0-4;11	5;0-5;11	adults
t'	144	157	133	95	196
t	126	154	125	92	199
t^h	141	141	120	82	199
k'	156	166	132	98	193
k	154	170	131	94	199
k^h	146	159	124	95	200

Table 3.2: The distribution of consonant tokens produced by Korean speaking children and adults that were used in the acoustic analysis (Database II).

3.3 Analysis method

3.3.1 Acoustic measures

In addition to temporal characteristics of phonation-type such as VOT, we examined the spectral characteristics of Korean stops such as *f0* and H1-H2 by inspecting the following vocalic onset.

3.3.1.1 VOT

The method of VOT measurement was identical to the description in Section 2.2.3. See Figure 2.2 and also Figure 3.1.

3.3.1.2 Fundamental frequency: *f0*

The *f0* was measured by taking the reciprocal of the interval between two neighboring pulses at 20 ms after the voicing onset, as shown in Figure 3.1. The glottal pulses were automatically detected using the pulse function in Praat (this is an autocorrelation-based periodicity detector). The analysis window was taken at 20ms after the voicing

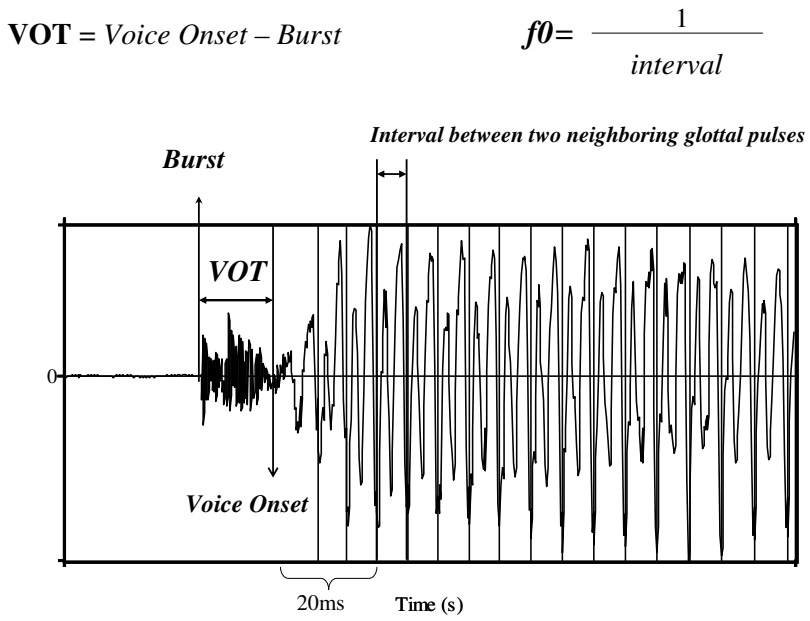

Figure 3.1: *f0* measurement.

onset instead of immediately at the voicing onset because it was observed that the function was more reliable (higher correlation coefficient) several glottal pulses after the exact onset of voicing.

3.3.1.3 Spectral tilt: H1-H2

We used H1-H2 as our spectral tilt measurement. As shown in Figure 3.2, it was measured by subtracting the amplitude (in dB) of the second harmonic from the amplitude of the first harmonic in the Fast Fourier Transform spectrum generated based on a 25ms long analysis Hamming window beginning at the voicing onset. The frequency location of the first harmonic was automatically detected by Praat referring

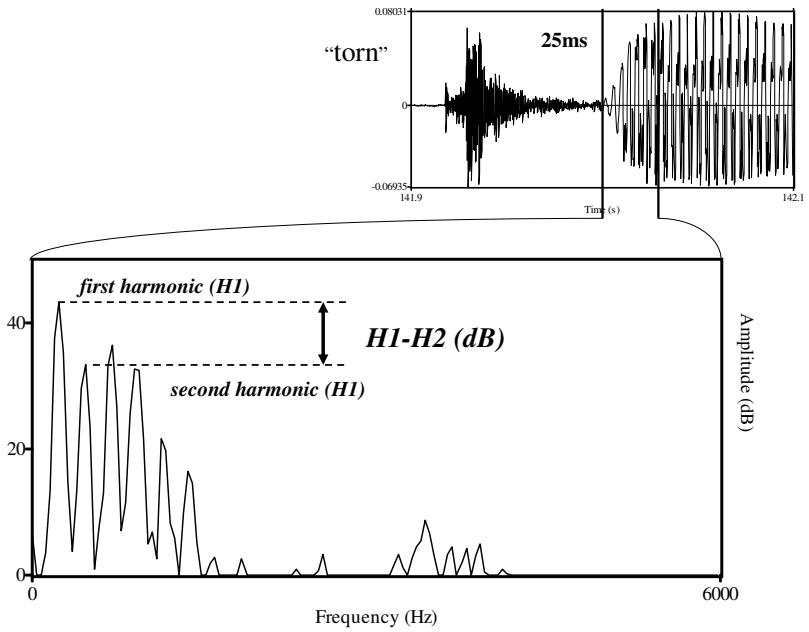

Figure 3.2: H1-H2 measurement.

to the fundamental frequency as an initial value, and then the researcher checked the precise frequency locations for the first and the second harmonic in a token by token manner in order not to be confused between DC noise and H1 or between H2 and A1 (amplitude of the first formant). There were instances where there was no clear second harmonic due to weak periodicity at the voicing onset. For these tokens, the frequency at twice the first harmonic was regarded as the frequency of the second harmonic, and the amplitude at that frequency location was taken as H2.

3.3.2 Statistical analysis

3.3.2.1 The mixed effects logistic regression model

Mixed effects logistic regression models are useful in at least three ways to estimate the relationship between the dependent variable and the predicting parameters. First, we can use such models to see whether each independent variable makes a significant contribution to the prediction of the dependent variable. That is, for each independent variable x_i, we can ask whether the associated coefficient β_i is equal to 0 or not equal to 0. Second, if all three parameters are significantly predictive, and the parameters are normalized to cover the same range, the absolute values of the coefficients can be compared directly to assess the relative size of the contribution. Third, we can compare the goodness of a model that includes a parameter to that of a model that does not include that parameter. Since the model produces the predicted category based on the coefficients of all the acoustic parameters, the percent match between the category predicted by the model and the category from the data can suggest how well the model performs in predicting the dependent variable using a given set of independent variables. For instance, we compare a model with VOT, *f0* and H1-H2 to a model with VOT parameter alone to assess the relative roles of VOT *f0* and H1-H2 in, say, in differentiating the tense stop from the other two types. If the prediction accuracy of the model with all three parameters is no better than the prediction accuracy of the VOT only model, VOT would be counted as the most essential acoustic parameter in characterizing the tense stop type from the non-tense categories.

The mixed effects logistic regression models for the adult productions are different from ones for the child productions in two ways. First, the adult models use the target categories (e.g., /t',k'/ versus non-/t',k'/) as the dependent variable, whereas

the child model uses the transcribed categories (e.g., [t',k'] versus non-[t',k']) as a dependent variable. Second, the adult models consider the speaker's gender as a factor, but the child models do not. The speaker's gender was included to control for sex differences (e.g., adult males have lower *f0* and generally less breathy voice quality than females due to the morphological changes in the larynx at puberty) and also to test whether any of the parameters significantly interact with the speaker's gender (e.g., men might have especially tense glottal settings as a marker of male identity).

The formulae for the adult models are given in Equation 3.1[1] and Equation 3.3 and the ones for the child models are given in Equation 3.2[2] and Equation 3.4. In the equations, the coefficients (i.e., β_1,β_2, β_3 ...) of each parameter indicate the size of the effect in determining the tenseness of the item, if the coefficient is a significantly effective variable in the model. The greater the absolute value of the coefficient, the more influential the variable it is. To remove the magnitude differences among measurement units (millisecond, decibel and hertz), three acoustic parameters were standardized using z-score transformation.

Using the formulae we have two equivalent sets of mixed effects logistic regression models. One set of mixed effects logistic regression models predicts the tense vs. the non-tense types (i.e., the lax and the aspirated types) as the dependent variable, and the other set of models predicts the lax vs. the aspirated types as the dependent variable.

[1]R.code:`lmer(target.category~logVOT.scale*factor(gender)+h1h2.scale*factor(gender)+f0.scale*factor(gender)+(logVOT.scale+h1h2.scale+f0.scale|Subject),data=adult.dat.stat,family=binomial)`

[2]R.code:`lmer(transcribed.category~logVOT.scale+h1h2.scale+f0.scale+(logVOT.scale+h1h2.scale+f0.scale|Subject),data=kid.dat.stat,family=binomial)`

$$(3.1)\quad log(\frac{/phonationType/}{1-/phonationType/}) = \beta_0 + \beta_1 logVOT + \beta_2 logVOT : Gen. + \beta_3 f0 + \beta_4 f0 : Gen. + \beta_5 H1H2 + \beta_6 H1H2 : Gen. + \gamma Speaker$$

where Gen. refers to the adult speaker's gender.

$$(3.2)\quad log(\frac{[phonationType]}{1-[phonationType]}) = \beta_0 + \beta_1 logVOT + \beta_2 f0 + \beta_3 H1H2 + \gamma Speaker$$

$$(3.3)\quad log(\frac{/phonationType/}{1-/phonationType/}) = \beta_0 + \beta_1 logVOT + \beta_2 logVOT : Gen. + \gamma Speaker$$

where Gen. refers to the adult speakers' gender.

$$(3.4)\quad log(\frac{[phonationType]}{1-[phonationType]}) = \beta_0 + \beta_1 logVOT + \gamma Speaker$$

3.4 Results

3.4.1 Adult productions

3.4.1.1 VOT, *f0* and H1-H2

Figure 3.3 shows the same adult VOT values as in Figure 2.3 but separated by gender (female vs. male). Vertical lines indicate the median values for the three types of Korean stops. Values for all three stops were in the lag VOT range. While the VOT range for the tense stops overlaps minimally with the ranges for the lax or aspirated

stops, lax stops share a wide range of VOT values with aspirated stops. This almost complete overlap in VOT values between lax and aspirated stops with no overlap between lax and tense stops seen in studies from the 1960s is consistent with the results of Silva (2006), (Wright, 2007), and Kang and Guion (2008), where the longer lax stop VOTs were identified as indicating a sound change in a progress. Younger Korean speakers born since the 1970s tend to have longer VOT values in the lax stops that are almost identical to their values for aspirated stops, whereas older speakers produce lax stop VOT values which are intermediate between those of the tense stops and those of the aspirated stops.

A gender-differentiated pattern of stop VOT distributions is also observed in Figure 3.3. Figure 3.3 shows that male speakers tend to have a significantly larger mean VOT difference between lax and aspirated stops than female speakers do. A repeated measure ANOVA showed a significant interaction between gender (male versus female) and phonation type (lax versus aspirated stops) (gender * phonation type) [$F(1,19)=5.9348$, $p<0.05$]. The greater overlap of VOT values between lax and aspirated stops by Korean female speakers may be related to the sound change to longer VOT in lax stops, while the better separation of VOT between lax and aspirated stops produced by Korean male speakers may indicate a somewhat more conservative form of shorter VOT in lax stops.

Figure 3.4 shows histograms for H1-H2, the breathiness measure. While tense stops have smaller H1-H2 values than lax and aspirated stops in both genders, H1-H2 values for all three types produced by Korean female speakers were greater than those produced by male speakers. The gender-related differences in H1-H2 are expected, given the morphological differences in the larynx between men and women such that adult males have a smaller proportion of the length of glottis taken up by space between the arytenoids as well as more massive vocal folds proper (Titze, 1989). In

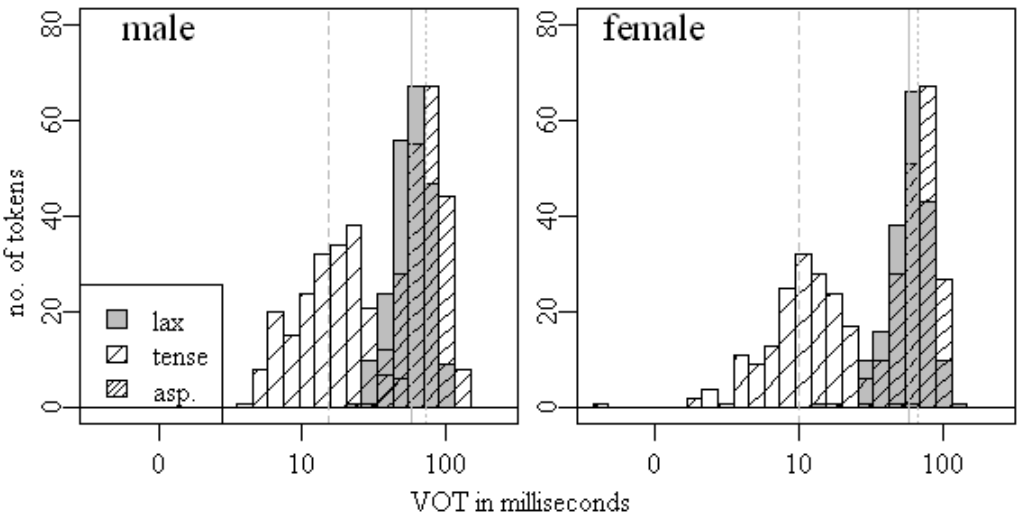

Figure 3.3: Histograms of VOT measured in lax, tense and aspirated stops produced by Korean male and female adults. The medians are indicated by vertical lines.

male speakers' H1-H2 distributions, the peak for the tense stop was distinguished from the peaks for the aspirated and lax stops in terms of a qualitative difference between negative vs. positive H1-H2 values. In contrast, the distinction between the peaks for the tense stop and the two other types in female speakers' H1-H2 distributions was mostly made by having smaller H1-H2 positive values for tense stops. Two repeated measures two-way ANOVA showed a significant interaction between gender and consonant type (tense vs. aspirated and tense vs. lax) with respect to the mean H1-H2 values: [$F(1,19)=19.204$, $p<0.001$] for tense vs. lax stops and [$F(1,19)=27.494$, $p<0.001$] for tense vs. aspirated stops. These gender-related patterns of H1-H2 differences suggest that Korean male speakers tend to produce the tense stop consonants with a more pressed glottal configuration than female speakers do.

It was along the *f0* dimension that lax stops were separated from the aspirated stops by having lower values. As shown in the histograms in the lefthand panels in

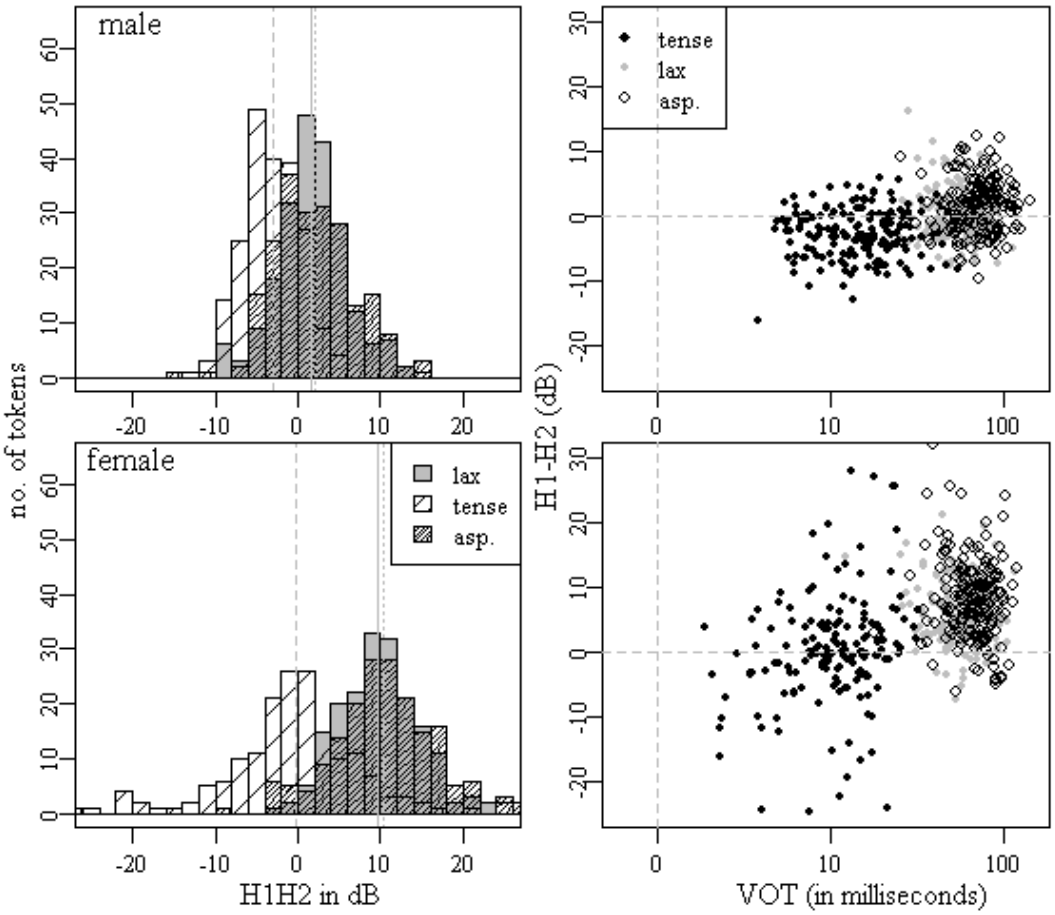

Figure 3.4: Histograms of H1-H2 measured in lax, tense and aspirated stops produced by Korean male and female adults speakers. The medians are indicated by vertical lines. The panels on the right are the scatterplots of H1-H2 as a function of log transformed VOT in milliseconds.

Figure 3.5, *f0* values at vowel onset after lax stops were lower than those after aspirated stops. The rightnad panels in Figure 3.5 also show the separation of the three types of Korean stops in VOT (log transformed VOT in milliseconds) scatterplots against *f0*. In both scatterplots, the *f0* values after tense stops were intermediate between the *f0* values for lax and aspirated stops and overlapped with both other distributions. In the females' pattern, there were a number of tense and aspirated stops

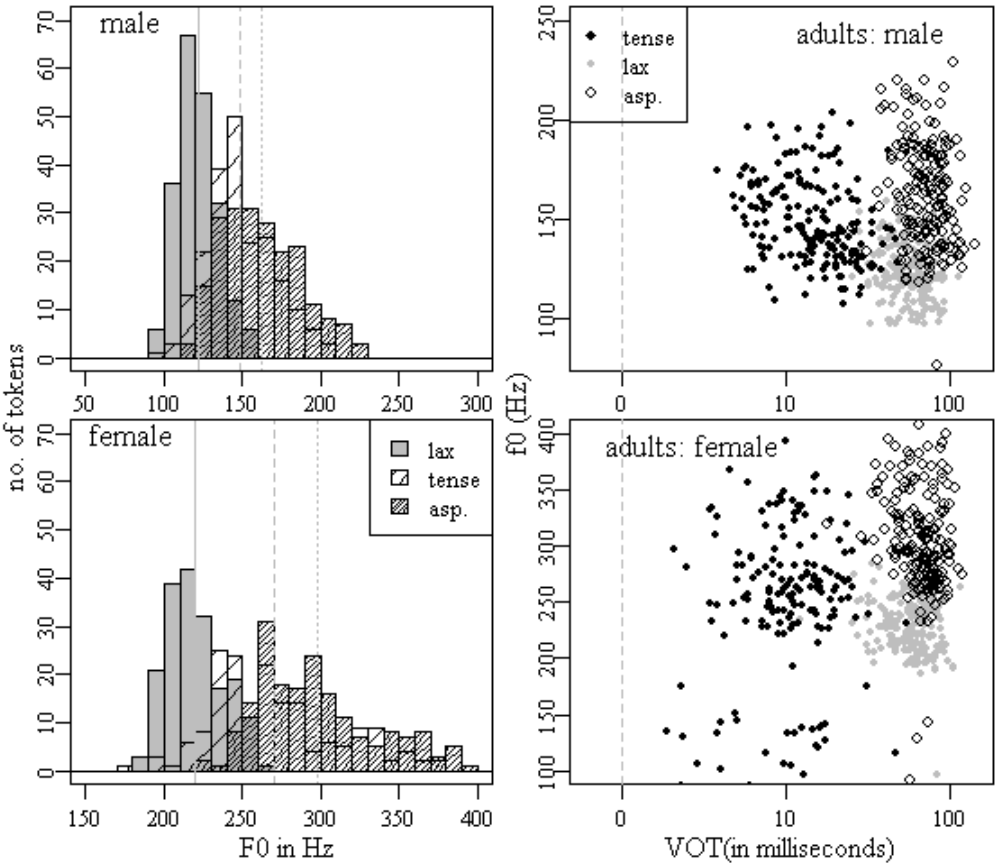

Figure 3.5: Histograms of *f0* measured in lax, tense and aspirated stops produced by Korean male and female speakers. The medians indicated by vertical lines. The panels on the right are the scatterplots of *f0* as a function of log transformed VOT in millisecond.

with low *f0* values under 150 Hz due to glottalization at the vocalic onset after the consonant. In Figure 3.5, the data points under 150 Hz are tokens with glottalization.

3.4.1.2 Mixed effects model of logistic regression

We built the mixed effects logistic regression models using Equation 3.1 and Equation 3.3. Since there are two possible predictions regarding mastery (see Section 2.4.1), we built two sets of models. In one set, we contrasted tense stops to the

other two types – i.e., predicted the target "tense" – and in the other, we contrasted aspirated stops to lax stops – i.e., predicted the target "aspirated", given non-tense.

Table 3.3 and Figure 3.6 show the results for the first set of models where the dependent variable was the categorical variable "tense". According to the fixed effect coefficients in Table 3.3, there was a significant main effect of VOT without a significant interaction with gender. The effect of H1-H2, however, did interact with gender. H1-H2 was significantly effective in the Korean male speakers' production, yet it was not significantly effective in the female speakers' production. The main effect of *f0* was not significant in the model. Figure 3.6 plots the probability curves of tense vs. nontense generated by the inverse logit function. The slopes of the curves were based on the coefficients of significantly effective parameters in each gender. Because there was no significant interaction between gender and VOT, the estimated probability curves for the male and female models are identical.

When the prediction accuracies were calculated to assess the goodness of this multi-parameter model based on the coefficients in Table 3.3, 97% of model predicted categories were identical to the target stop phonation provided by the data. When the other model of VOT as the only predictor(Equation 3.3) was implemented, there was the same 97% of correct prediction in the adult speakers' model. The goodness of the model with three acoustic parameters was no better than the goodness of the model with the VOT parameter alone.

Table 3.4 and Figure 3.7 show the results of the second set of mixed effects logistic regression models, where the dependent variable was aspirated vs. lax stops. In these models, the main effects of VOT (log-transformed) and *f0* were significant in both genders in differentiating the aspirated stop from the lax stop. However, H1-H2 did not have a significant main effect in doing so. The effect of *f0* was significantly different between male and female productions due to a significant interaction between

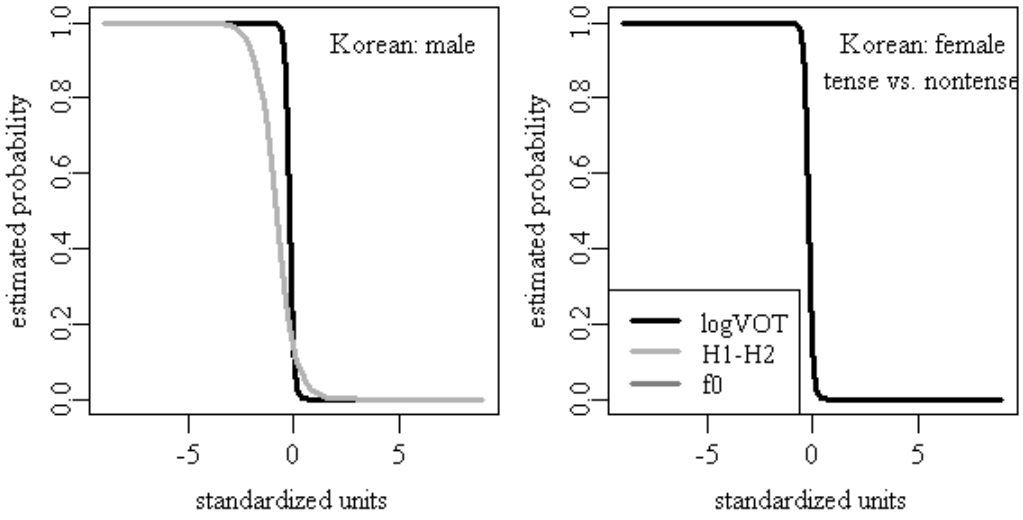

Figure 3.6: The probability curves of tense vs. nontense estimated from the mixed effects models of logistic regressions in Korean (Equation 3.1). The estimated probability of '1' indicates 'tense', whereas '0' indicates 'non-tense'. The curves were generated by the inverse logit function only for the significant effects. The exact values of the coefficients are shown in Table 3.3.

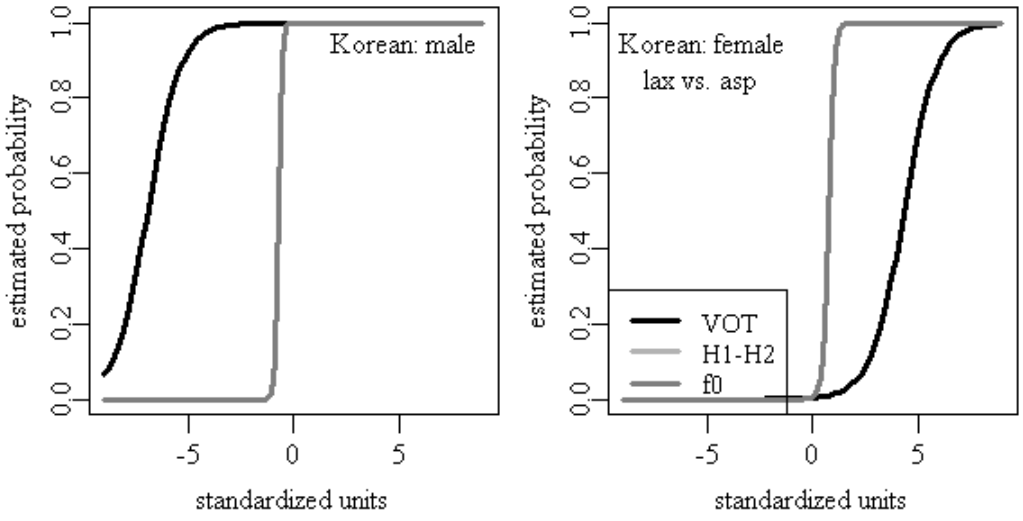

Figure 3.7: The probability curves of lax vs. asp. estimated from the mixed effects models of logistic regressions in Korean (Equation 3.1). The estimated probability of '1' indicates 'asp.', whereas '0' indicates 'lax'. The curves were generated by the inverse logit function only for the significant effects. The exact values of the coefficients are shown in Table 3.4.

Random effects:	subj				
Name	Variance	Std.Dev.	Corr		
(Intercept)	1.31960	1.1487			
logVOT	6.31280	2.5125	-0.195		
h1h2	0.27562	0.5250	-0.694	0.842	
f0	0.81831	0.9046	-0.801	-0.432	0.124
Number of obs:	1059,	groups: subj,	20		

Fixed effects:					
Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-2.1981	0.4923	-4.465	7.99e-06 ***	
logVOT	-10.1284	1.2547	-8.072	6.89e-16 ***	
h1h2:factor(gender)m	-2.2578	0.6482	-3.483	0.000496 ***	
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 3.3: Table of the coefficients of the mixed effects logistic regression model (Equation 3.1) in Korean adult speaker's productions (tense vs. non-tense). Only the significant independent variables ($p<0.05$) are listed in the fixed effects table. Since the base gender of the model was female ('f'), the significant interaction between gender and independent variables are indicated with 'm' (the male speakers) next to the interaction term.

gender and *f0* ($p<0.05$), while there was no significant interaction between gender with VOT. According to the coefficients of the two effective parameters in the model ('Estimate' in Table 3.4), *f0* was more influential in the differentiation between the aspirated stops and the lax stops in both genders. A larger effect size of *f0* over VOT is shown as a steeper slope in the probability curves in Figure 3.7.

To assess the goodness of the model with multiple acoustic parameters, we compared the categories predicted by the model with the categories in the data. The full model matched the target stop phonation-type for 91% of the tokens. Compared to this prediction accuracy of the model with VOT and *f0* as predicting parameters, the VOT-only model could only predict 63% of the tokens correctly. When the *f0* was

Random effects:	subj				
Name	Variance	Std.Dev.	Corr		
logVOT	0.17981	0.42404	0.019		
h1h2	0.27691	0.52622	0.483	-0.764	
f0	4.67287	2.16168	-0.848	-0.018	-0.406
Number of obs:	715,	groups: subj,	20		

Fixed effects:					
Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-5.5978	1.0684	-5.240	1.61e-07 ***	
logVOT	1.2857	0.2338	5.499	3.82e-08 ***	
f0	7.5635	1.1238	6.730	1.69e-11 ***	
factor(gender)m	15.1693	1.7808	8.518	< 2e-16 ***	
f0:factor(gender)m	4.7457	1.8465	2.570	0.0102 *	
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 3.4: Table of the coefficients of the two mixed effects models of logistic regression (Equation 3.1) in Korean adult speaker's productions (lax vs. aspirated) separated by the speaker's gender.

the only predicting parameter in the mixed effects model, 85% of the model prediction was correct in the adult speakers' model.

To summarize, the acoustic characteristics of three different types of Korean stop phonation showed that tense stops in Korean were discriminated from the other two stops predominately by VOT whereas lax and aspirated stops were differentiated from each other mostly by *f0* values.

It is noted that this particular *f0* measure was not effective in differentiating the tense stop from the lax and aspirated stops in Korean adult speaker's stop productions. In differentiating tense stop from either lax or aspirated stops, the model with VOT alone as an independent parameter predicted the target consonant type as well as the model with more parameters (VOT and H1-H2) did. This suggests that

VOT is a sufficient acoustic parameter in differentiating tense stops from the other two types of stop phonation in Korean.

In contrast, the models with *f0* performed better in differentiating the aspirated from the lax stops. There was a great improvement of prediction accuracy from the model with the VOT parameter alone to the model with VOT and *f0*, suggesting that VOT alone is not a sufficient acoustic parameter in the distinction between the lax and aspirated stops in Korean.

3.4.2 Child productions

3.4.2.1 VOT, *f0* and H1-H2

Figure 3.8 shows the VOT distributions of Korean speaking children's stop productions. The VOT distributions are plotted according to their intended target phonation-type categories, separated by gender and into age group at one-year intervals.

In the youngest age group (2;0-2;11) shown in the top panels, the VOT values of tense stops were concentrated at a short lag range with relatively sharp peaks skewed toward zero VOT. By contrast, the lax and aspirated stops had relatively large variability in VOT values, covering both the short and the long lag ranges. In the older children's productions (3;0-5;11), the lax and aspirated stops had distributional peaks at longer VOT values which were clearly separated from the peaks for tense stops. At all ages, the medians for the lax stops were not greatly different from those for the aspirated stops, which is similar to the pattern in Korean adult speakers' VOT distributions.

Among the three different phonation types, the VOT values of the tense stops were most adult-like in the two year olds' productions in that they were realized as

short lag VOT values despite wider variability. However, unlike adults' tense stops, there were a small number of tense stops that were made with lead VOT values in Korean-speaking children's productions. The duration of the prevoicing lead in children's tense stops was relatively short compared to the lead VOT for voiced stops of other languages studied in Lisker and Abramson (1964), which ranged from -170 ms to -45 ms in Spanish, for instance.

Figure 3.9 shows the distributions of H1-H2 for the intended targets of tense, lax and aspirated stops, separated by gender and by age group. In the two year olds' H1-H2 distributions (top panels), there was no distinction among the three types of stops. Values for all three types were almost completely overlapped with each other. In the older children's productions, the H1-H2 values showed some separation between the tense stops and the other two types. The H1-H2 values of tense stops were generally lower than those of lax and aspirated stops. In addition, although the median H1-H2 values of all three types of stops were positive, tense stops had the most with negative H1-H2 tokens.

Figure 3.10 displays the distributions of *f0* values for tense, lax and aspirated stops produced by Korean children separated by age group and gender. While there was some overlap among the three types of stops along the *f0* range, the lax stops were mostly distributed at a lower *f0* range and the tense and aspirated stops were distributed at a higher *f0* range where they were almost completely overlapped with each other. This was true even for two year olds' stop productions.

3.4.2.2 Transcribed categories and acoustic parameters

Figure 3.11 and Figure 3.12 show the children's productions again, but this time plot the phonation types identified by the transcriber, as a function of each acoustic parameter. The two figures correspond to the two sets of regression models. In the first

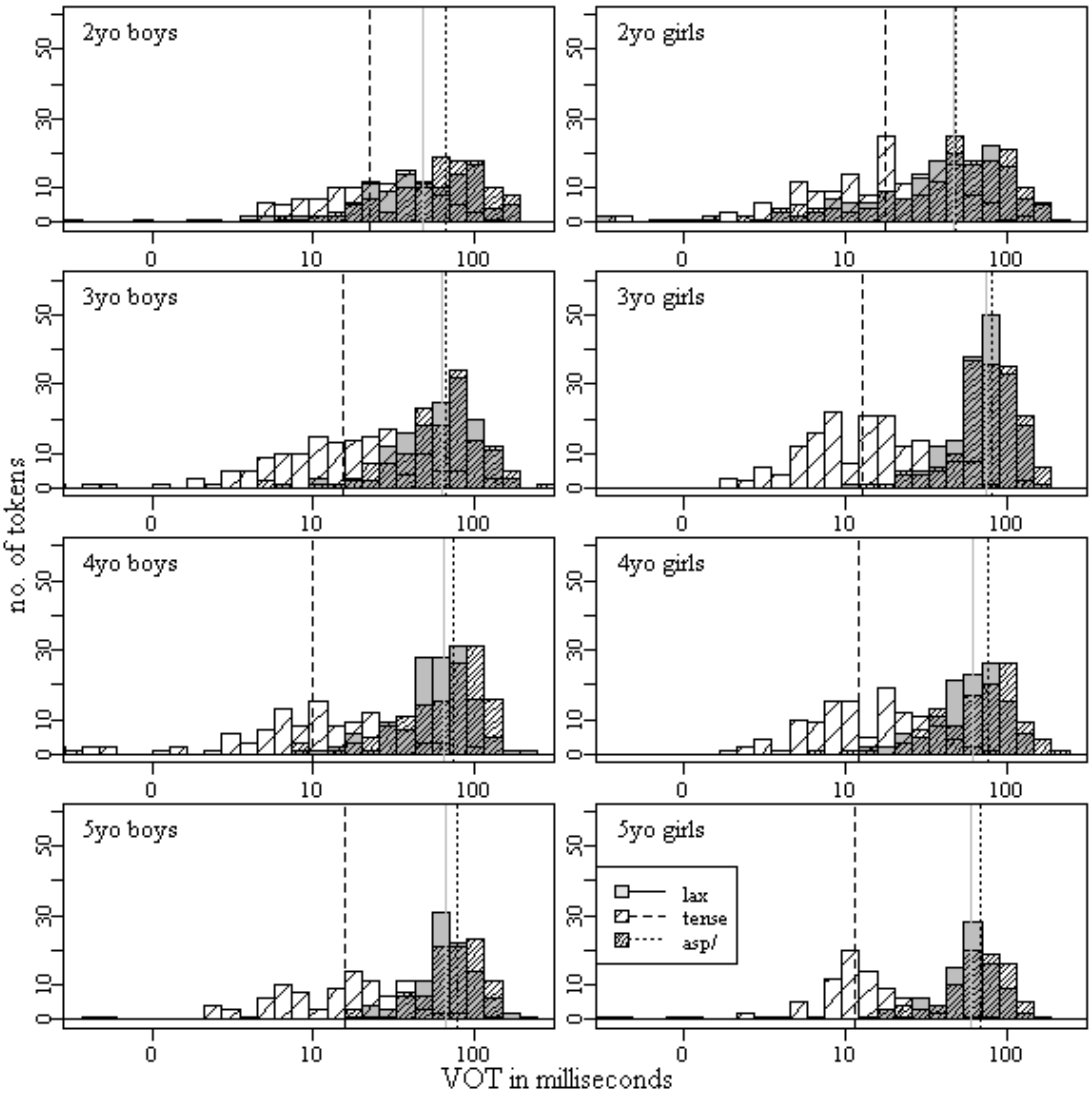

Figure 3.8: Histograms of VOT in three phonation-types of Korean stop produced by children from ages 2;0 to 6;0. The place of articulation distinction has been collapsed. Vertical lines indicate VOT medians of tense, lax and aspirated stops.

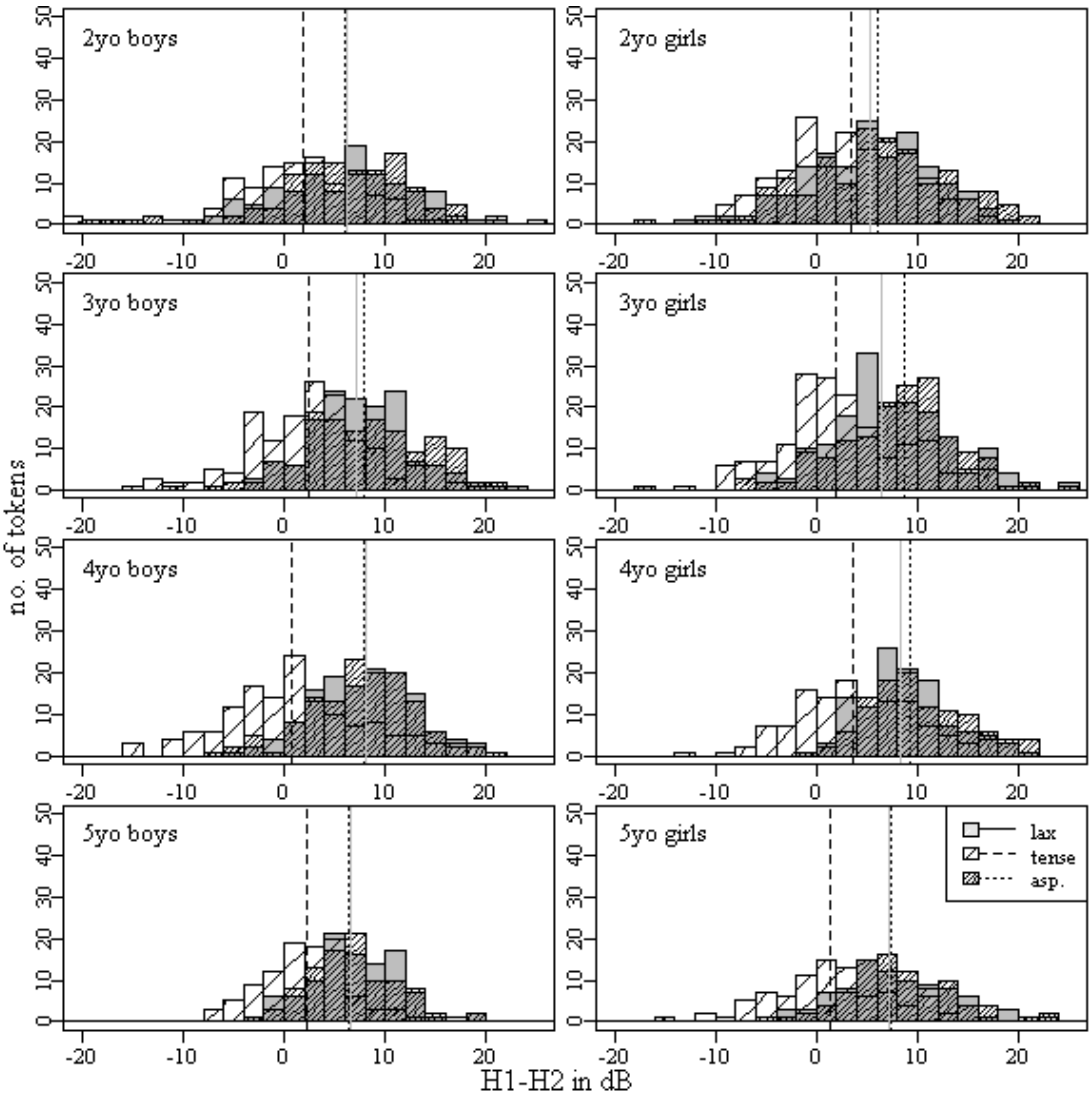

Figure 3.9: Histograms of H1-H2 in three phonation-types of Korean stop produced by children from ages 2;0 to 6;0. The place of articulation distinction has been collapsed. Vertical lines indicate the VOT medians of tense, lax and aspirated stops.

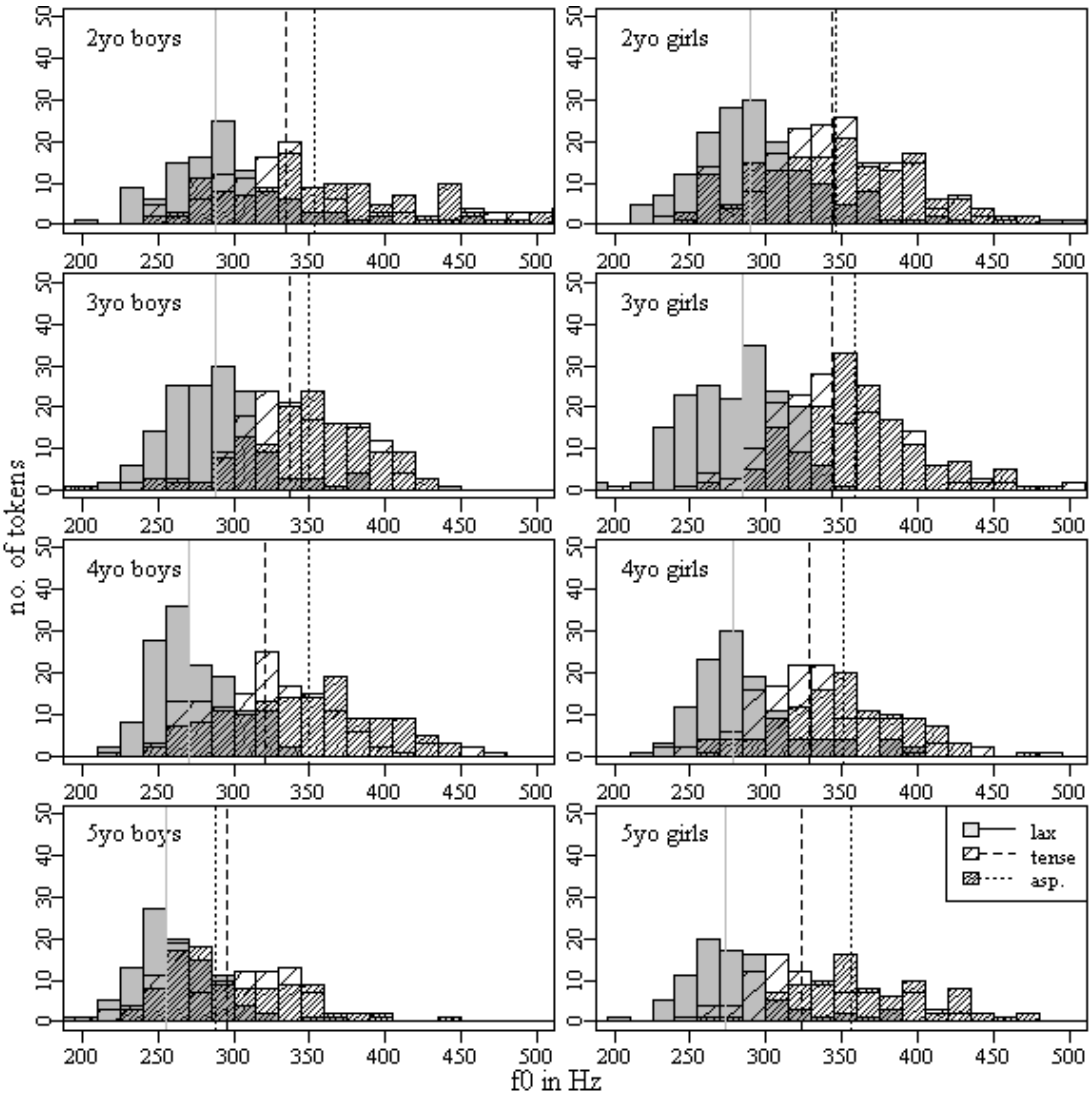

Figure 3.10: Histograms of *f0* in three phonation-types of Korean stop produced by children from ages 2;0 to 5;11. The place of articulation distinction has been collapsed. Vertical lines indicate the VOT medians of tense, lax and aspirated stops.

set, the models predict whether the transcriber identified the token as tense or non-tense (i.e., lax and aspirated types). Figure 3.11 accordingly shows the proportion of transcribed tense categories for each individual child's production that showed particular acoustic values for VOT, *f0* and H1-H2. The leftmost panel in Figure 3.11 shows that the transcriber's identification of the tense type was negatively correlated with VOT – i.e., the higher the VOT values, the less likely the transcriber was to identify the token as the tense type. The *f0* and H1-H2 were also correlated with the transcriber's judgment of tense type versus non-tense type in a way that tokens of higher *f0* values and tokens of lower H1-H2 (i.e., less breathier quality) were likely to be identified as the tense type. The inverse logit curves were overlaid on the scatterplots to capture the overall trend of tense proportions over the total productions.

In the second set of models, the children's productions that were identified as non-tense by the transcriber are sub-divided into those identified as aspirated versus those identified as lax to see how the judgment of the aspirated type over the lax type is related to the acoustic parameters. Figure 3.12 shows the proportion of tokens transcribed as the aspirated stop among each child's tokens that were identified as non-tense stop. as a function of the binned acoustic values of VOT, *f0* and H1-H2. There was a trend such that the transcriber's judgment of the aspirated category was associated with longer VOT values. Similarly, higher *f0* values of the tokens predicted a higher proportion of aspirated type judgments by the transcriber. By contrast, H1-H2 was not at all predictive regarding the judgment of the aspirated type over the lax type. The proportion of tokens judged as the aspirated stop did not change as the values of H1-H2 increased.

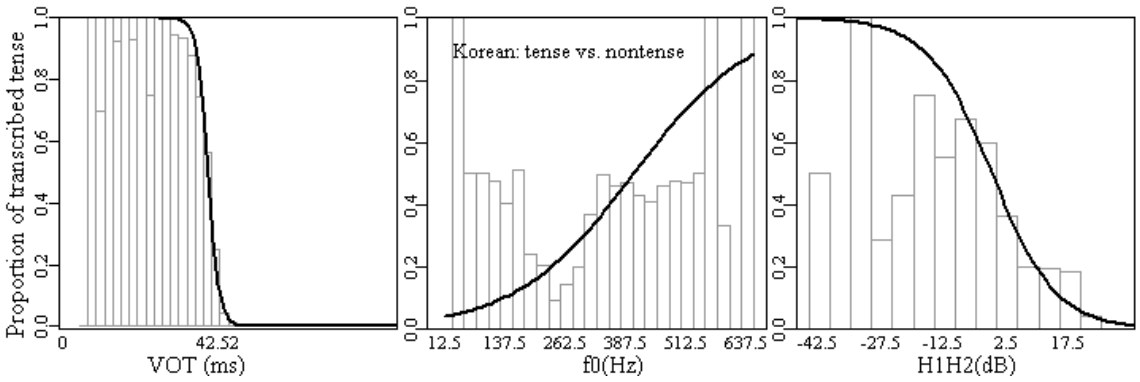

Figure 3.11: The proportion of tense stops (vs. non-tense stops), identified by the transcriber as a function of VOT, *f0* and H1-H2. The bar heights indicates the proportion of transcribed tense type in the binned acoustic values. The curves were overlaid to capture the trend, which were generated by the inverse logit of mixed effects logistic regression.

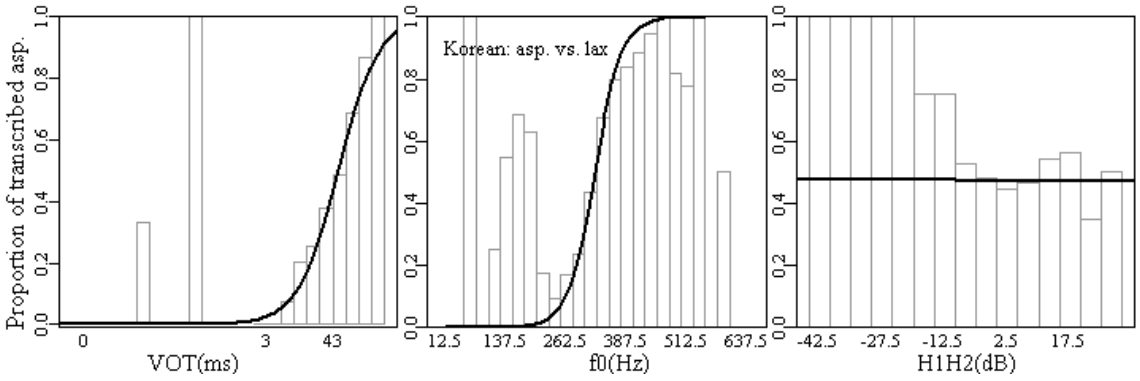

Figure 3.12: The proportion of aspirated stops (vs. lax stops), identified by the transcriber as a function of VOT, *f0* and H1-H2. The bar heights indicates the proportion of transcribed aspirated type in the binned acoustic values. The curves were overlaid to capture the trend, which were generated by the inverse logit of mixed effects logistic regression.

3.4.2.3 Mixed effects model of logistic regression

Table 3.5 and Table 3.6 show the output of the regression models exploring the relationships between the acoustic parameters and the transcriber's judgment of the phonation type, which were foreshadowed in Figure 3.11 and Figure 3.12, respectively. That is, as we did with the adult targets we constructed mixed effects logistic regression models as described in Equation 3.2 and Equation 3.4 in Section 3.3.2.

In the first set, the dependent variable was the transcriber's judgment of tense vs. non-tense (i.e., lax and aspirated) types. Table 3.5 provides the coefficients of each independent parameter that was significant in differentiating the tense category from the non-tense types in the full model, which used Equation 3.2. VOT, H1-H2 and *f0* were all significantly effective parameters in this model. Among them, VOT contributed most. The coefficient for VOT was 6.29 times greater than that for *f0*, and 6.23 times greater than that for H1-H2. This larger effect of VOT over H1-H2 and *f0* is shown as a steeper slope in the probability curve in Figure 3.11 and Figure 3.13 (Recall that the values of the three acoustic parameters were standardized for a comparison across the coefficients of parameters. Figure 3.13 replots the curves in Figure 3.11 using the common standardized units on the x-axis, so that the curves can be overlaid and directly compared.)

When we evaluated the goodness of the model by comparing the categories predicted by the model to the categories provided by the transcriber, correct predictions were made in 86% of the data. Since we are interested in how dominant the VOT parameter is in determining the tense stops from the non-tense stops, we made another model based on Equation 3.4 that predicts the dependent variable only based on VOT. This VOT model again could correctly predict 86% of the data. This might suggest that VOT is a sufficient acoustic parameter that determines the tense

Random effects:	subj				
Name	Variance	Std.Dev.	Corr		
(Intercept)	0.48864	0.69903			
logVOT	2.93200	1.71231	-0.090		
f0	0.16793	0.40979	0.105	-0.032	
h1h2	0.18709	0.43254	-0.254	-0.095	0.040
Number of obs:	3111,	groups: subj,	67		
Fixed effects:					
Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-0.53300	0.11233	-4.745	2.09e-06 ***	
logVOT	-3.75383	0.26035	-14.418	< 2e-16 ***	
f0	0.59679	0.09146	6.525	6.81e-11 ***	
h1h2	-0.60254	0.09177	-6.566	5.18e-11 ***	
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 3.5: Table of the coefficients of the two mixed effects models of logistic regression in Korean child speakers' productions (tense vs. non-tense)

category in contrast with lax and aspirated stops in the transcription of children's productions.

The second set of mixed effects logistic regression models used the transcriber's judgment of aspirated vs. lax types of phonation in children's stop productions in all the age groups. In other words, this one was applied just to the subset of tokens that the transcriber identified as either lax or aspirated and the dependent variable was the identification of the token as aspirated. Again, standardized scores for log-transformed VOT, for H1-H2 and for *f0* were the independent parameters and the speaker differences were taken as the random effect of the model.

Table 3.6 shows the coefficients of each independent parameter in the model and Figure 3.14 repeats the two significantly predictive ones on a common x-axis, so that their slopes can be compared directly. Only VOT and *f0* (but not H1-H2)

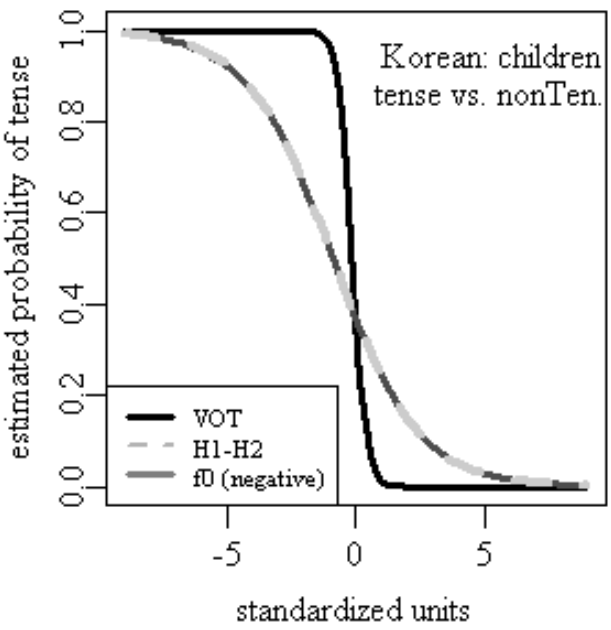

Figure 3.13: The probability curves of tense vs. non-tense estimated from the mixed effects logistic regression models in Korean children's productions. The estimated probability of '1' indicates 'tense', whereas '0' indicates 'non-tense'. The exact values of the coefficients are shown in Table 3.5. The direction of *f0* slope is made negative for a direct comparison of slopes across parameters.

were significant in differentiating the aspirated from the lax stops. Between the two significantly effective parameters, *f0* was more influential than VOT as indicated by its larger coefficient and steeper slope. The *f0* parameter was 1.99 times effective in differentiating the stops transcribed as aspirated from the stops transcribed as lax.

The goodness of the model was assessed by comparing the phonation type categories predicted by the model and the categories identified by the transcriber. Only 79% of tokens matched. When another model was made with log-transformed VOT as the only predicting parameter (Equation 3.4), 64% of the tokens matched. The percent match was higher, 77%, in the model where *f0* was the only independent variable. The prediction accuracy differences among the models suggest that the

Random effects:	subj				
Name	Variance	Std.Dev.	Corr		
(Intercept)	0.96495	0.98232			
logVOT	0.57670	0.75941	-0.456		
f0	2.90107	1.70325	0.058	-0.082	
h1h2	0.17099	0.41351	0.395	0.294	0.265
Number of obs:	1629,	groups: subj,	56		

Fixed effects:					
Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-0.2980	0.1596	-1.867	0.0619 .	
logVOT	1.3730	0.1600	8.580	<2e-16 ***	
f0	2.7427	0.2682	10.225	<2e-16 ***	
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 3.6: Table of the coefficients of the two mixed effects models of logistic regression in Korean child speakers' productions (asp. vs. lax)

differentiation of aspirated stops from lax stops might not be sufficiently made by VOT alone, but that *f0* needs to be taken into a consideration as well.

3.5 Summary

We examined the acoustic characteristics of three different Korean stops of the phonation-type contrast produced by adults and children with three goals in mind. First, we wanted to describe how adults realize the contrast between tense and non-tense and between aspirated and lax. Second, we wanted to see when children begin to make adult-like patterns. Third, we wanted to know whether/how the production accuracy of lax, tense and aspirated stops is based on the acoustic characteristics. We were interested in the relatively early mastery of tense stops, whose articulatory characteristics are known to be complex with respect to the supraglottal aerodynamic

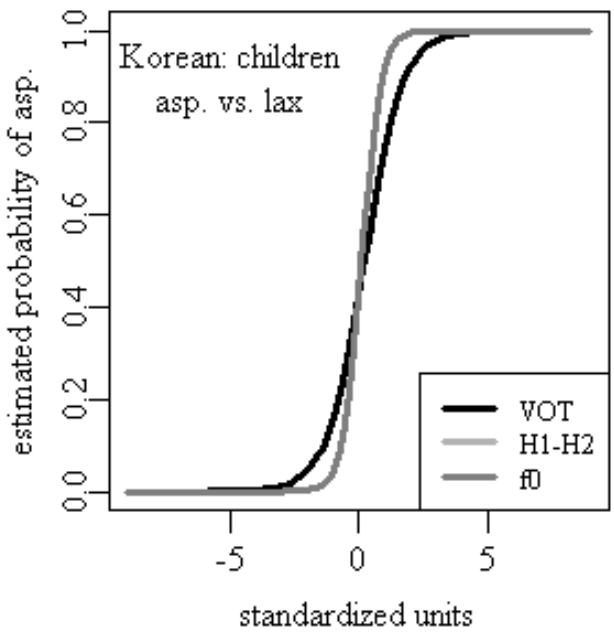

Figure 3.14: The probability curves of lax vs. asp. estimated from the mixed effects models of logistic regressions in Korean children's productions. The estimated probability of '1' indicates 'asp.', whereas '0' indicates 'lax'. The exact values of the coefficients are shown in Table 3.6.

conditions and the intrinsic laryngeal muscle activities. Given that the acoustic dimension of VOT cannot discriminate all three types of stop phonations, we were motivated to examine various acoustic dimensions other than VOT in order to investigate how the acoustic characteristics of children's stop productions affect the judgment of production accuracy.

Our logistic regression models for the adult patterns suggested that VOT differentiates tense from non-tense but it cannot distinguish aspirated from lax stops. The distinction between the lax and the aspirated stops was instead made by the *f0* dimension. The lax stops have a lower *f0* than the aspirated stops. These acoustic characteristics support the idea that there has been a recent diachronic change in the

role of VOT in the distinction betweem, the three types of stops, as suggested by Silva (2006) and Kang and Guion (2008).

The early mastery of tense stops as gauged by the native-speaker transcription might be explained by the role of the VOT cue used in the transcription judgment and the acoustic properties of children's stop productions. Less affected by other acoustic aspects such as *f0* or H1-H2 in judging the accuracy of tense stops, the transcriber will identify the children's productions of the tense target dominantly by the VOT values of the tokens. Children produce a short lag VOT value before mastering a long lag VOT or a lead VOT. Having a short lag VOT suffices for the transcriber to identify it as a correct production of a tense stop. The dominant role of VOT in encoding the contrast between tense and other phonation types and the emergence of *f0* as the dominant cue differentiating the aspirated stops form the lax stops, together seem to explain the high production accuracy in Korean children's production. Given the role of VOT in the transcription analysis, the relatively high production accuracy of the Korean tense stop is not such an exceptional pattern after all.

CHAPTER 4

VOICED AND VOICELESS STOPS IN JAPANESE AND ENGLISH

This chapter discusses the Japanese stop voicing contrast and compares the efficacy of VOT as an acoustic parameter for predicting voicing accuracy judgments for the Japanese children's productions relative to its predictive power for the English speaking children's productions. In the analysis described in Chapter 2, Japanese voiced stops were mastered later than their voiceless counterparts and even later than English aspirated stops. Figure 2.4 and Figure 2.5 showed much lower accuracy for the Japanese voiced stops produced by two- and three-year olds in Database I, and Figure 2.5 showed crossover points for 75% transcribed accuracy for coronal and dorsal voiced stops in Database II at 43 months and at 50 months as compared to well before 24 months for the voiceless stops. Such a pattern of later mastery for voiced stops is what we would predict for a language with a "true" voicing contrast, as Japanese has been described to be.

However, the Japanese pattern is puzzling if the actual contemporary VOT distributions are examined. As Figure 2.3 showed, the Japanese voiced stops did not necessarily have lead VOT values. A majority of Japanese female speakers' voiced stops were found to have short lag VOT values. This VOT characteristic of Japanese voiced stops might predict an early mastery of voiced stops because a short lag VOT is the typical value in young children's productions across languages.

On the other hand, if we consider these voiced stops in relation to the contrasting voiceless stops, the predictions become less straightforward. According to Figure 2.3, the Japanese voiceless stops were realized with a lag VOT value that is intermediate between the short lag and the long lag VOT values observed in English adult speakers' productions of voiced and voiceless stops. It has never been investigated how children might master a contrast between voiced stops that can have lead or short lag VOT values and stops that are necessarily voiceless but realized with VOT values in an intermediate range between "true" unaspirated and "true" aspirated types.

The fact that Japanese children's voiced stops are judged as incorrect despite the adult-like short lag VOT values leads to the suspicion that Japanese children's voiced stops require other acoustic aspects besides short VOT to be accepted as adult-like voiced stops. The goal of this chapter is to compare the Japanese and English patterns in order to determine whether another acoustic dimension can help in predicting how Japanese children's stops will be transcribed.

Specifically, we explore the acoustic properties of breathiness and fundamental frequency as other possible acoustic dimensions that could characterize the Japanese stop voicing contrast with a short lag VOT. We chose to explore the dimension of voice quality and fundamental frequency inspired by the findings of Shimizu (1989) and Takada (2004a). Shimizu (1989) demonstrated that voiceless stops tend to be followed by a significantly higher *f0* than that of voiced stops in Japanese. He also observed the pattern of rising *f0* curves made from the immediate vocalic onset through a 60ms region in the voiced stop. Takada (2004a) showed that the glottal configuration of a voiced stop is adducted at the time of the burst regardless of whether the stop shows a short lag VOT or a lead VOT. This adducted glottal configuration of the voiced stop is expected to result in a more modal voice quality as compared to a breathy

voice quality associated with a more open glottal configuration (Hanson, 1997). As we did for the Korean productions, we quantify this modal voice versus breathy voice by measuring the amplitude difference between the first harmonic and the second harmonic (H1-H2).

To evaluate the potential differences between the Japanese voiced vs. voiceless contrast and the English contrast, we estimate the relative contributions of the three acoustic parameters VOT, H1-H2 and *f0* to the voicing category judgment using the same kind of mixed effects logistic regression models described in Chapter 3. The hypothesis that VOT alone will not sufficiently differentiate Japanese voiced stops from voiceless stops suggests that the model will identify a smaller contribution from VOT to the prediction of the voiced stop category in Japanese than in English. Conversely, the relative contribution of one or both of the other acoustic parameters (H1-H2 and *f0*) to the prediction of voicing contrast in the Japanese model should be greater than in the English model. In this chapter, we compare the relative contributions of acoustic parameters to the prediction of children's stop voicing accuracy between English and Japanese so that we can relate the puzzling pattern of lower production accuracy in Japanese voiced stops to the language-specific use of acoustic parameters in distinguishing the two contrasting voicing categories.

4.1 Method

4.1.1 Materials, subjects and tasks

The English and Japanese subsets of Database II were analyzed in this chapter. The descriptions of the materials, subjects and task were provided in Chapter 2. The target words with the stops in a high-vowel context (/i/ and /u/) were excluded in the acoustic analysis, since the amplitude boost to the second harmonic from the

	Japanese		English	
age group	girls	boys	girls	boys
2;0-2;11	6	7	7	6
3;0-3;11	7	5	4	7
4;0-4;11	6	8	6	6
5;0-5;11	5	7	6	8
adults	10	10	7	8

Table 4.1: The age distributions of child subjects speaking Japanese and English in Database II.

language	target	2;0-2;11	3;0-3;11	4;0-4;11	5;0-5;11	adults (m.)	adults (f.)
Japanese	d	90	99	121	106	89	90
	t	97	96	121	108	89	90
	g	107	100	123	102	89	90
	k	104	91	118	104	89	90
English	d	69	91	106	118	63	72
	t	71	80	105	117	63	72
	g	34	32	36	40	21	24
	k	36	28	36	37	21	24

Table 4.2: The distribution of the consonant tokens produced by Japanese or English speaking children and adults that were used in the acoustic analysis (the subsets of Database II).

low first formant in a high vowel makes it more difficult to interpret H1-H2 as a breathiness index. Table 4.1 shows the age distribution of child and adult speakers used in the acoustic analysis. Table 4.2 shows the number of tokens analyzed for each target consonant.

4.2 Analysis method

4.2.1 Acoustic measures

We used three acoustic measures to compare the voicing contrasts in Japanese and English: VOT, H1-H2 and *f0*. The measurement methods are described in Chapter 2 and Chapter 3.

4.2.2 Statistical analysis

Mixed effects logistic regression models were applied to the adult and child stop productions of the Japanese and English subsets. For adult productions, the models predicted whether a token is a voiceless stop target (dependent variable). For child productions, the dependent variable is not the target but the transcribed voicing category (i.e., transcribed as [t] or not [t] for Japanese and [t^h] or not [t^h] for English). In both cases, there was a complete model that includes as independent variables VOT (log-scale), H1-H2 and *f0*. We formulate another smaller model that has a single independent variable of VOT.

The speaker's language (English vs. Japanese) was a between-subjects factor in the models for both adults and children, with speaker gender included in the adult models. The units of each acoustic variables were standardized for a comparison of

coefficients across the variables. The models for the adults are given in Equation 4.1[1] and Equation 4.3[2], and the models for children are in Equation 4.2[3] and Equation 4.4[4].

(4.1)

$$log(\frac{/voiceless/}{1-/voiceless/}) = \beta_0 + \beta_1\mathbf{logVOT} + \beta_2\mathbf{logVOT} : \mathbf{Lg.} + \beta_3\mathbf{logVOT} : \mathbf{Gen.}$$
$$+\beta_4\mathbf{logVOT} : \mathbf{Gen.} : \mathbf{Lg.} + \beta_5\mathbf{f0} + +\beta_6\mathbf{f0} : \mathbf{Lg.} + \beta_7\mathbf{f0} : \mathbf{Gen.}$$
$$+\beta_8\mathbf{f0} : \mathbf{Gen.} : \mathbf{Lg.} + \beta_9\mathbf{H1H2} + \beta_{10}\mathbf{H1H2} : \mathbf{Lg.} + \beta_{11}\mathbf{H1H2} : \mathbf{Gen.}$$
$$+\beta_{12}\mathbf{H1H2} : \mathbf{Lg.} : \mathbf{Gen.} + \beta_{13}\mathbf{Lg.} + \beta_{14}\mathbf{Gen.} + \gamma\mathbf{Speaker}$$

where **Lg.** and **Gen.** refer to the speakers' language and gender, respectively.

(4.2)

$$log(\frac{[voiceless]}{1-[voiceless]}) = \beta_0 + \beta_1\mathbf{logVOT} + \beta_2\mathbf{logVOT} : \mathbf{Lg.} + \beta_3\mathbf{f0} + \beta_4\mathbf{f0} : \mathbf{Lg.}$$
$$+\beta_5\mathbf{H1H2} + \beta_6\mathbf{H1H2} : \mathbf{Lg.} + \beta_7\mathbf{Lg.} + \beta_8\mathbf{Gen.}$$
$$+\beta_9\mathbf{Lg.} + \beta_{10}\mathbf{Gen.} + \gamma\mathbf{Speaker}$$

[1] R.code:`lmer(target.category~logVOT.scale*factor(language)*factor(gender)+h1h2.scale*factor(language)*factor(gender)+f0.scale*factor(language)*factor(gender)+(logVOT.scale+h1h2.scale+f0.scale|Subject),data=dat.adult.stat,family=binomial)`

[2] R.code:`lmer(target.category~logVOT.scale*factor(language)*factor(gender)+(logVOT.scale|Subject),data=dat.adult.stat,family=binomial)`

[3] R.code:`lmer(transcribed.category~logVOT.scale*factor(language)+h1h2.scale*factor(language)+f0.scale*factor(language)+(logVOT.scale+h1h2.scale+f0.scale|Subject),data=dat.kid.stat,family=binomial)`

[4] R.code:`lmer(transcribed.category~logVOT.scale*factor(language)+(logVOT.scale|Subject),data=dat.kid.stat,family=binomial)`

(4.3)

$$log(\frac{/voiceless/}{1-/voiceless/}) = \beta_0 + \beta_1\mathbf{logVOT} + \beta_2\mathbf{logVOT} : \mathbf{Lg.} + \beta_3\mathbf{logVOT} : \mathbf{Gen.} + \beta_4\mathbf{logVOT} : \mathbf{Gen.} : \mathbf{Lg.} + \beta_5\mathbf{Lg.} + \beta_6\mathbf{Gen.} + \gamma\mathbf{Speaker}$$

where **Lg.** and **Gen.** refer to the speakers' language and gender, respectively.

(4.4)

$$log(\frac{[voiceless]}{1-[voiceless]}) = \beta_0 + \beta_1\mathbf{logVOT} + \beta_2\mathbf{logVOT} : \mathbf{Lg.} + \beta_3\mathbf{Lg.} + \beta_4\mathbf{Gen.} + \gamma\mathbf{Speaker}$$

4.3 Results

4.3.1 Adult productions

4.3.1.1 VOT, *f0* and H1-H2

The VOT distributions of English and Japanese adult stop productions plotted in Figure 2.3 are plotted again in Figure 4.1 but this time showing only the subset of tokens with non-high vowels and also separated by gender. The place distinctions are collapsed in both languages this time, because there was no statistically significant interaction of consonant place or stop voicing category according to a repeated measures two-way ANOVA.

In English, the VOT values for voiced vs. voiceless stops show almost no overlap, confirming that VOT is a successful acoustic parameter for the English stop phonation-type distinction. Voiced stops have a bimodal distribution with one peak

in the lead VOT region and the other in short lag VOT region. This pattern of distribution was found in both genders. As in English, Japanese voiced stops have two variants, a voiceless unaspirated variant (with short lag VOT) in addition to 'true' voiced stops (with lead VOT). This variation in Japanese is associated with gender; all the prevoiced tokens in female's productions were made by only two female speakers, whereas only two male speakers produced a majority of voiced targets with short lag VOT. In a repeated measures two-way ANOVA, there was a significant interaction between gender and phonation-type for the VOT parameter [$F(1,18)=12.811$, $p<0.005$]. VOT values for Japanese voiceless stops were different from those for English ones, in that Japanese voiceless stops have VOT values intermediate between those of English aspirated stops and unaspirated stops. There was found no clear-cut boundary between voiced stops and voiceless stops in Japanese along the VOT continuum when productions by all speakers were pooled together. This wide range of overlap between voiced and voiceless stops in Japanese reflects both the intermediate VOT values of voiceless stops (3ms-94ms) and the presence of voiceless variants of voiced stops (0ms-56ms).

Figure 4.2 shows the distributions of the breathiness measure values for English speakers. H1-H2 in English showed two clearly separated peaks for the voiced stops versus the voiceless stops. While English H1-H2 values for female speakers were higher than those for male speakers, H1-H2 values for voiceless stops in both genders were greater than those for voiced stops.

The bottom panels of Figure 4.2 show the distributions of H1-H2 produced by Japanese male and female speakers. Although there was more overlap than in English, H1-H2 values for voiceless stops were generally higher than those for voiced stops in both genders.

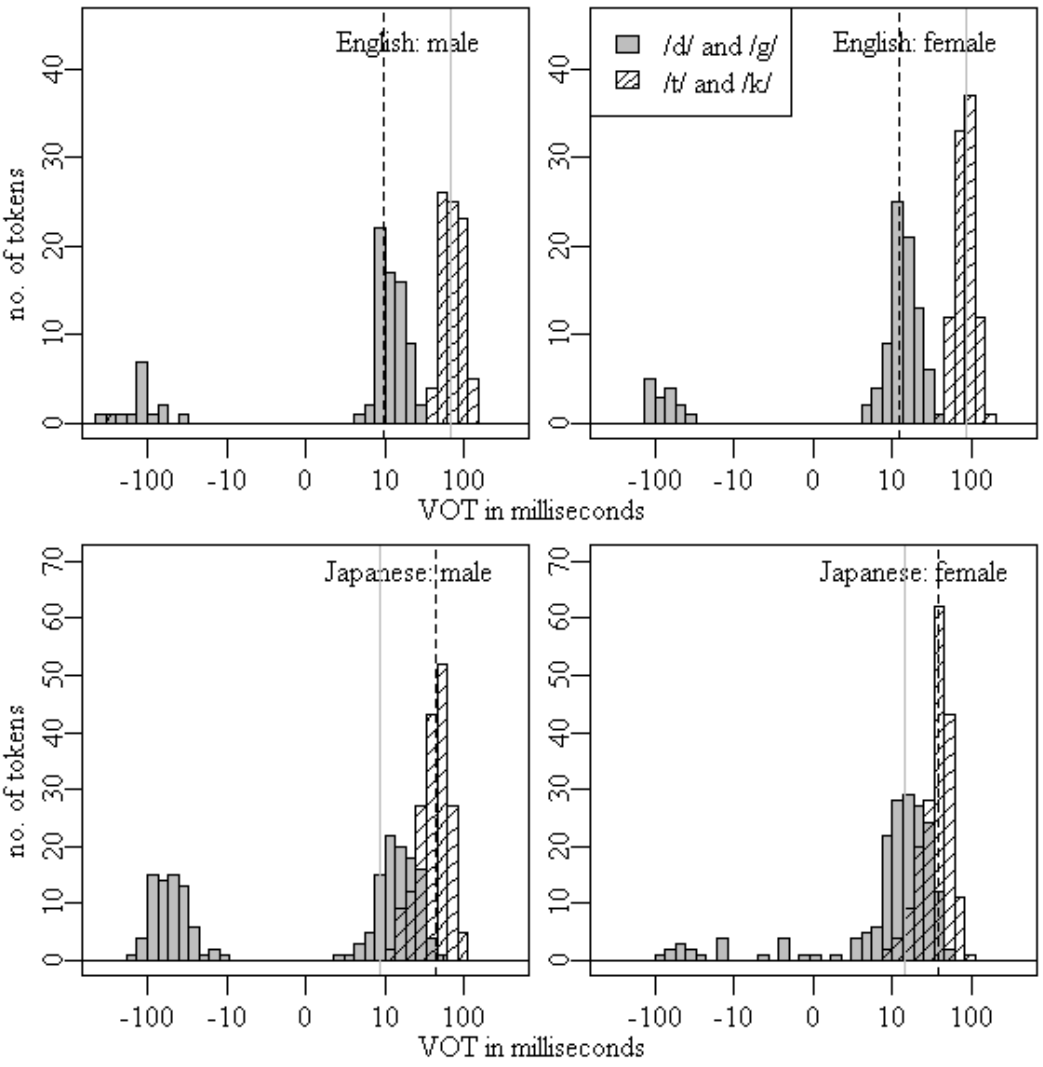

Figure 4.1: Histograms of English and Japanese stop VOTs in milliseconds produced by adult speakers. Vertical lines indicate the median values of VOT for each category.

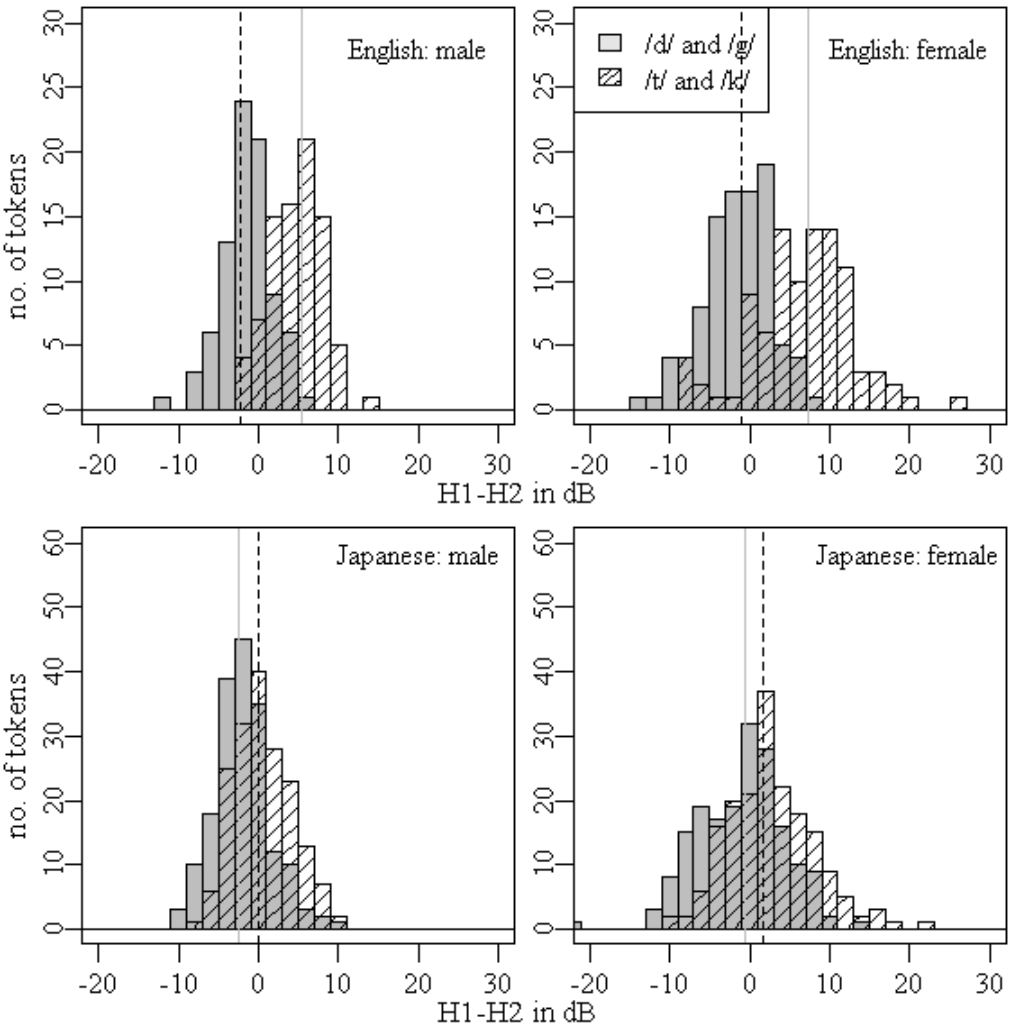

Figure 4.2: Histograms of H1-H2 measured in aspirated and unaspirated stops produced by English and Japanese adults. Vertical lines indicate the median H1-H2 values for each consonant voicing category.

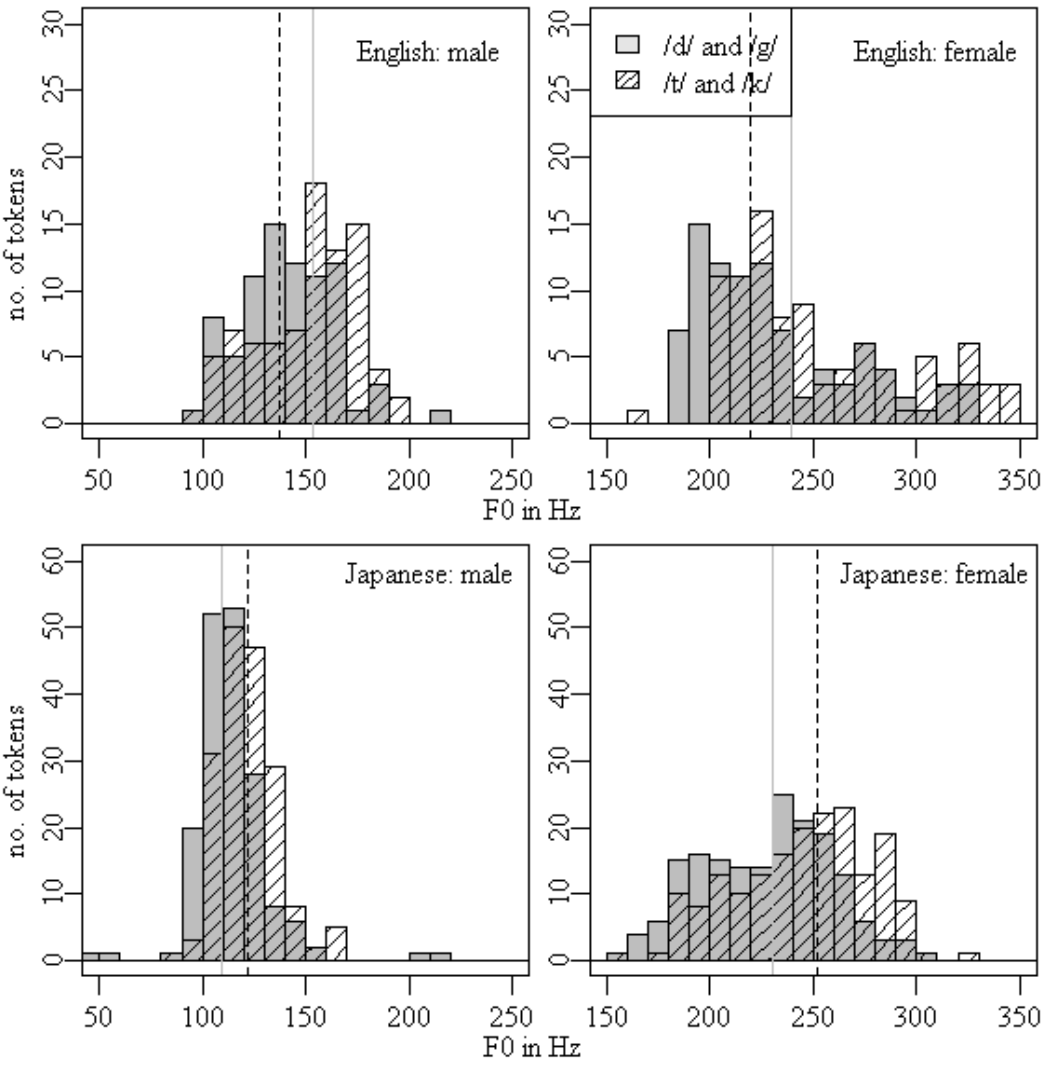

Figure 4.3: Histograms of English and Japanese stop VOT produced by adult speakers. Vertical lines indicate the median values of *f0* for each category.

Figure 4.3 shows histograms of the *f0* values 20 ms after vowel onset for each stop voicing category in English and Japanese, separated by gender. A common pattern across panels was that voiceless stops in both languages tended to have higher *f0* values than voiced stops did. This tendency was true for each gender. A significant interaction between gender and stop voicing category was found in both languages according to the repeated measures two-way ANOVA ([$F(1, 18)= 10.537$, $p<0.005$] in Japanese, [$F(1, 13)= 4.1833$, $p<0.1$] in English).

4.3.1.2 Mixed effects logistic regression model

Table 4.3 and Figure 4.4 show the results of the mixed effects logistic regression model where the target voiceless stops are predicted by VOT alone (Equation 4.3).

There was a significant interaction between language and VOT ($p<0.05$), and a significant interaction for the Japanese tokens between gender and VOT. The coefficient for English (38.7) was larger than either coefficient for Japanese (38.7-30.5 = 8.2 , 38.7-29.4=9.3), as indicated graphically by the steeper shape of the curve in Figure 4.4. When this mixed effects model of logistic regression was used to predict the target voiceless stops, the predicted output of the model matched the target category of voiceless stops in 99.16% of English productions, and 83.36% of Japanese productions. The differences in the effect size and the prediction accuracy suggest that the effectivness of VOT in English is greater than its effectiveness in Japanese in differentiating the target voiceless stops from voiced stops.

Table 4.4 and Figure 4.5 show the results for the mixed effects logistic regression model that included H1-H2 and *f0* as well as VOT (Equation 4.1). There was a significant main effect of VOT, but no significant language and VOT interaction, which is indicated by the equally steep slope of probability curves in the leftmost panel of Figure 4.5. The main effect of H1-H2 was also significant in the model with

Fixed effects:

Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-16.892	6.372	-2.651	0.00802	**
logVOT	38.728	13.836	2.799	0.00512	**
logVOT:lgGd(jpnF)	-30.541	13.863	-2.203	0.02759	*
logVOT:lgGd(jpnM)	-29.462	13.876	-2.123	0.03374	*
lgGd(jpnF)	14.720	6.378	2.308	0.02101	*
lgGd(jpnM)	14.626	6.381	2.292	0.02190	*
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 4.3: Table of the coefficients of the VOT estimated from the mixed effects model of logistic regression (Equation 4.3) that were significant at $p < 0.05$. The 'lgGd' factor refers to the token groups sorted by language and gender (English, Japanese female, and Japanese male).

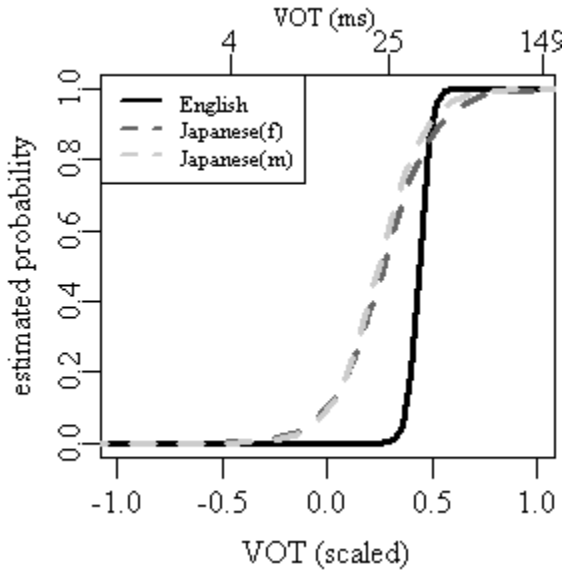

Figure 4.4: The curves of estimated probability with respect to log-transformed VOT. The estimated probability of '1' indicates the voiceless stop, whereas '0' indicates the voiced stop. The exact coefficient values are in Table 4.3.

Fixed effects:

Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-4.66700	0.48785	-9.566	< 2e-16	***
logVOT	10.84063	0.76041	14.256	< 2e-16	***
h1h2	1.09752	0.17067	6.431	1.27e-10	***
f0	-0.05399	0.37653	-0.143	0.886	
lgGd(jpnF):f0.scale	2.54368	0.53994	4.711	2.46e-06	***
lgGd(jpnM)	2.25899	0.52059	4.339	1.43e-05	***
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 4.4: Table of the coefficients of the mixed effects logistic regression model in English and Japanese adult speaker's productions (Equation 4.1). Non-high vowel context only.

no interaction between H1-H2 and gender. The middle panel of Figure 4.5 reflects this result by the slopes that are equally steep. The significant main effect of language in Japanese male speakers' tokens is shown as a separation, along the x-axis, of the Japanese male speakers' curve from the English and Japanese female speakers' (completely overlapped) curves. The effect of *f0* was significant only in Japanese female speaker's tokens.

4.3.2 Child productions

4.3.2.1 VOT

Figure 4.6 shows histograms of the VOT values for target voiced and voiceless stops produced by English speaking children, separated by gender and four different age groups. While the VOT distributions were highly adult-like in that the voiced stops were clustered at the short lag VOT range and the voiceless at the long lag range,

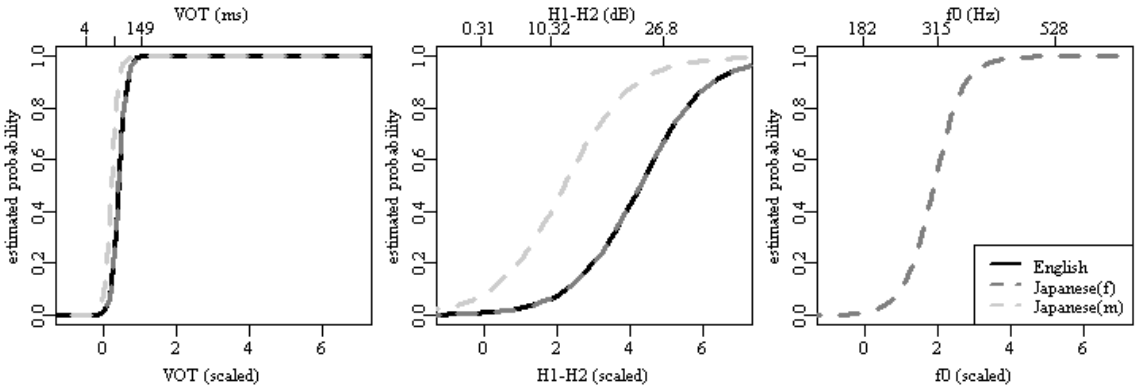

Figure 4.5: The curves of estimated probability with respect to VOT, H1-H2 and *f0*. The estimated probability of '1' indicates the voiceless stop, whereas '0' indicates the voiced stops. The coefficient values are shown in Table 4.4.

the youngest children showed more variability. That is, in the graphs for the English-speaking two year olds (top panels), some voiced stops are realized with long lag VOT values and some voiceless stops are realized with very long lag VOT values, resulting in a less peaky distribution, and for the girls, a somewhat greater median VOT value for the voiceless stops.

Figure 4.7 shows the VOT values for target voiced and voiceless stops produced by Japanese children across four different age groups. The patterns were separated by gender. A majority of the target voiced stops were realized with short lag VOT (The boys did not show the gender-related adult male pattern of producing the lead VOT variant). The voiceless stops had only slightly longer VOT values overall, but there were different degrees of overlap between the target voiced and voiceless stops across the four age groups. Older children appeared to make a better separation of distributional peaks between target voiced stops and voiceless stop, and the distance between medians for voiced and voiceless stops was somewhat larger for the five year olds than for the younger children (more comparable to the adult pattern).

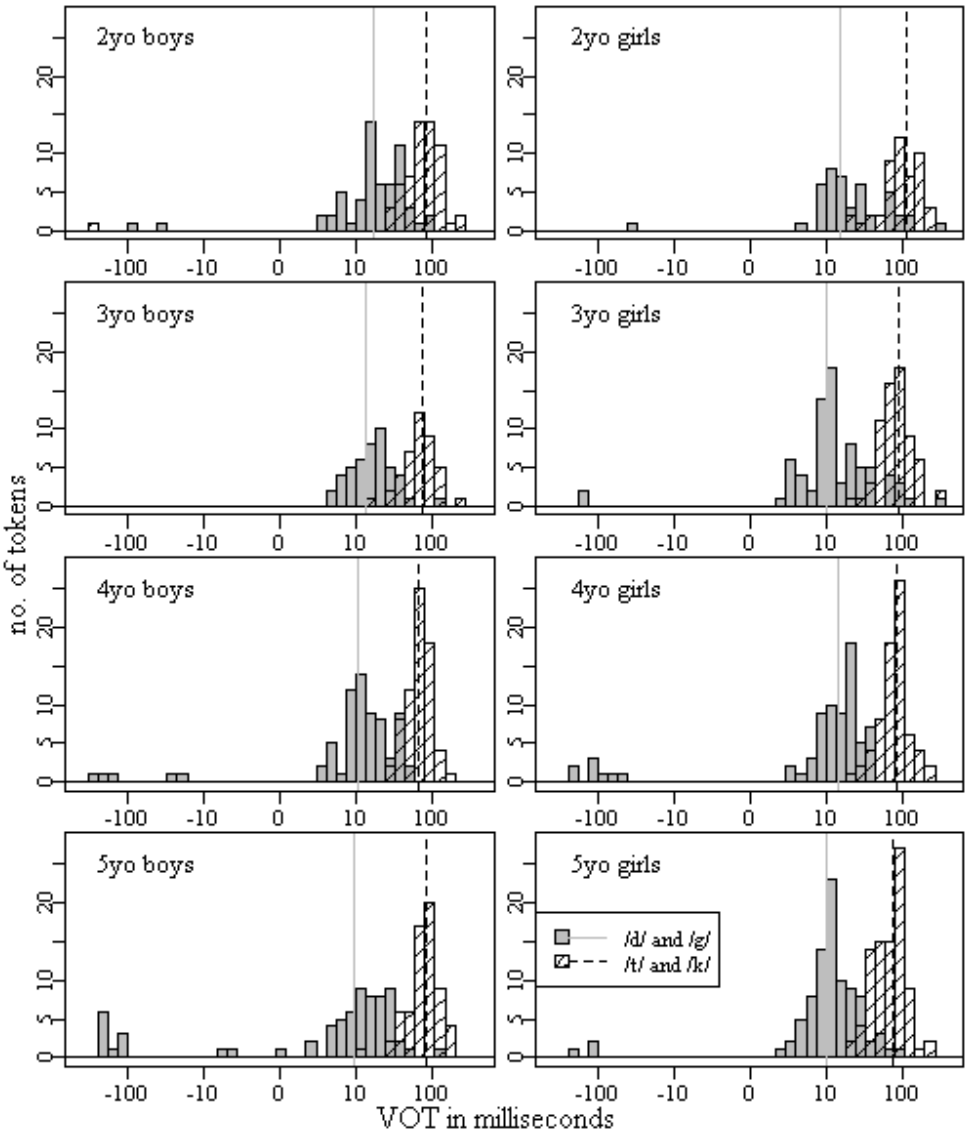

Figure 4.6: Histograms of English stop VOTs separated by the four different age groups (2;0-2;11, 3;0-3;11, 4;0-4;11 and 5;0-5;11) and gender (boys and girls). Vertical lines indicate the median VOT for voiced/voiceless stops in each age group.

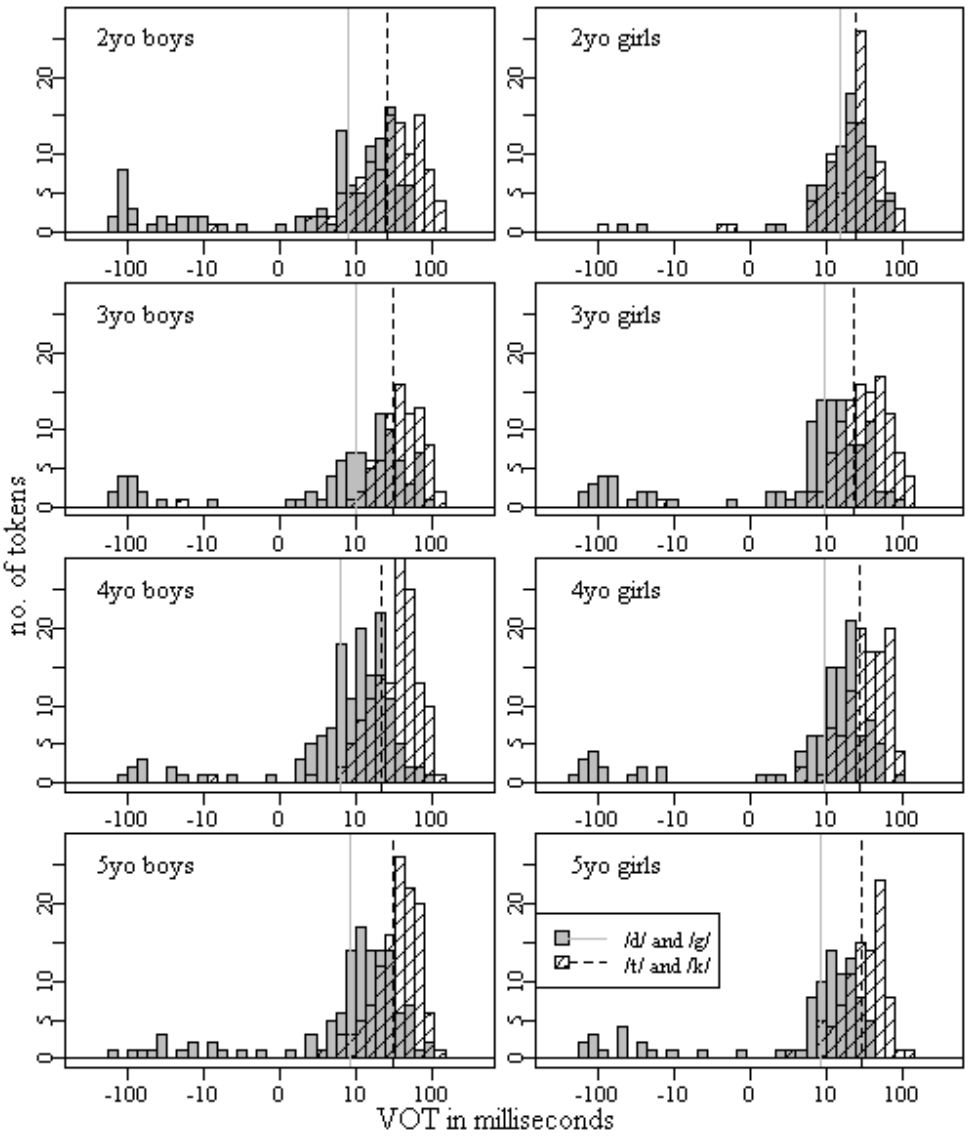

Figure 4.7: Histograms of Japanese stop VOT values separated by the four age groups (2;0-2;11, 3;0-3;11, 4;0-4;11 and 5;0-5;11) and gender (boys and girls). The VOT medians for each stop voicing category are indicated by vertical lines.

4.3.2.2 Transcribed categories and acoustic parameters

Figure 4.8 shows the proportion of tokens that were transcribed as voiceless as a function of each of the acoustic values of VOT, H1-H2 and *f0*. The three panels in the top row are for the English-speaking children's stops and the three panels in the bottom row are for the Japanese-speaking children's stops. In each panel, the heights of the bars indicate the proportion of tokens that were transcribed as voiceless. The overlaid curves show the results of the mixed effects logistic model that included all three parameters as independent variables (Equation 4.2). These curves are replotted in the three panels of Figure 4.10 where the x-axis shows the "scaled"(z-score) units that were used in the model.

Table 4.6 and Figure 4.10 show the significant coefficients from that model and Table 4.5 and Figure 4.9 show the results of the simpler mixed effects logistic regression that included only VOT as a parameter (Equation 4.4). In the simpler model, there was a significant interaction between language and VOT ($p< 0.05$). The interaction coefficient was greater for English than for Japanese, suggesting that the transcriber for the English-speaking children was influenced more by the VOT values of the tokens than was the transcriber for the Japanese speaking children. This greater effect of VOT for English than for Japanese is reflected in the steeper slope of the English curve than the Japanese curve. The mixed effects logistic regression with a single VOT variable could predict the transcribed voiceless stops correctly in 93.02% of the English-speaking children's productions as compared to 80.11% of the Japanese-speaking children's stops.

When H1-H2 and *f0* were added to VOT in the more complex model (Equation 4.2), there was again a significant interaction of language with VOT but only significant main effects for the other two parameters. However, the coefficients for

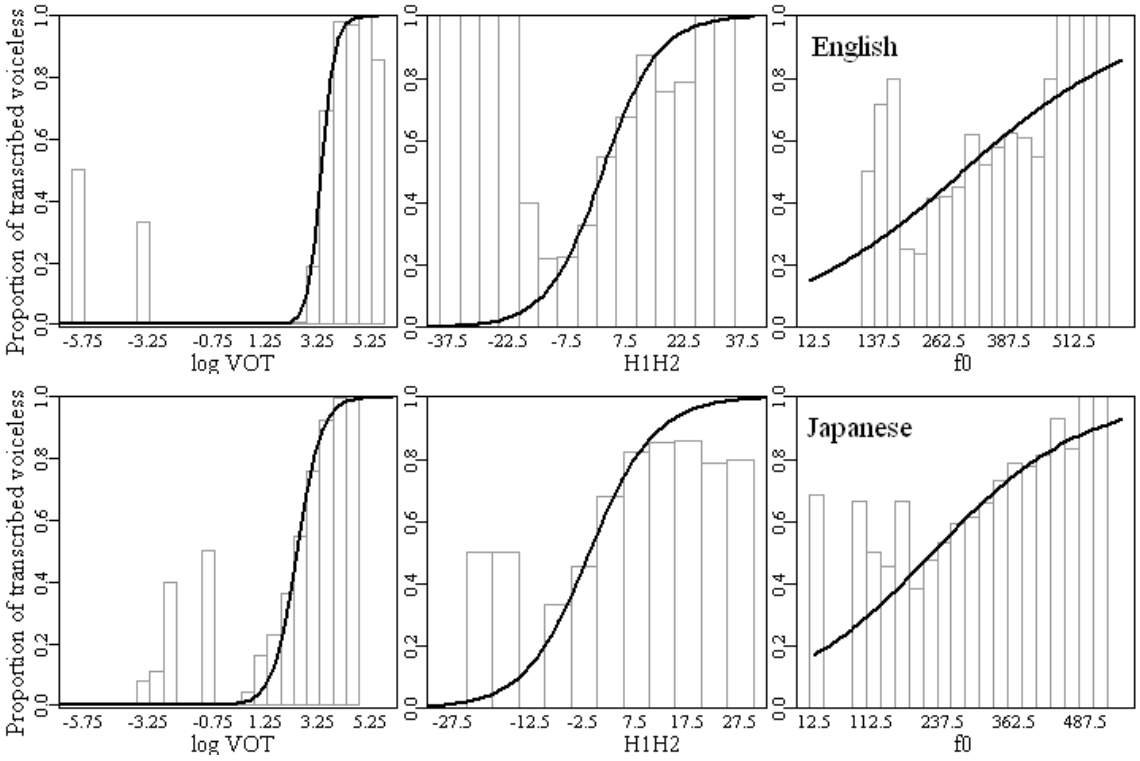

Figure 4.8: The proportion of voiceless stops (vs. voiced stops) identified by the native speaker transcribers as a function of VOT, *f0* and H1-H2. The curves were overlaid to capture the trend, which were generated by the inverse logit of coefficients from the mixed effects logistic regression.

Fixed effects:

Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-2.7767	0.2351	-11.812	<2e-16	***
logVOT	7.8613	0.5176	15.189	<2e-16	***
lg(j)	3.3656	0.2442	13.784	<2e-16	***
logVOT:lg(j)	-4.5386	0.5519	-8.223	<2e-16	***
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 4.5: Table of the coefficients of the two mixed effects models of logistic regression based on the transcribed voiced vs. voiceless stops produced by English and Japanese children's productions. The model has the log-transformed VOT as the only predicter variable. Non-high vowel context only.

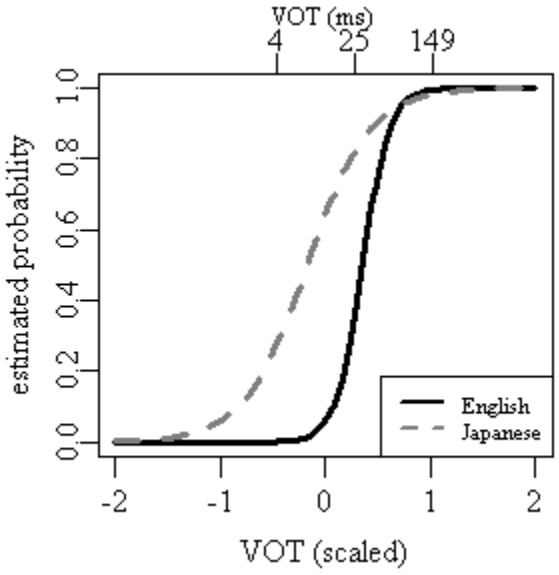

Figure 4.9: The curves of estimated probability with respect to VOT parameter generated from the mixed effects logistic regression model of English- and Japanese-speaking children's production. The estimated probability of '1' indicates the transcribed voiceless stops, whereas '0' indicates the transcribed voiced stops. The exact coefficient values are shown in Table 4.5

VOT were greatest among the three parameters in both languages. Because the interaction coefficient for VOT with English was more than twice that for VOT with Japanese (as indicated by the steeper slope for English in the left most panel of Figure 4.10), the contribution of the other two parameters was still relatively greater in Japanese. Conversely, we can say that the relative effect of VOT compared to the effects of H1-H2 and *f0* was higher in the English tokens than in the Japanese tokens. VOT was 13.12 times (7.86/0.598) and 13.21 times (7.86/0.594) more effective than H1-H2 and *f0*, respectively, in predicting the transcription of voiceless stops produced by English-speaking children, whereas VOT was 5.37 times ((7.86-4.64)/0.598) and 5.40 times ((7.86-4.64)/0.594) more effective than H1-H2 and *f0* in predicting the transcriptions of the voiceless stops produced by Japanese-speaking children.

Fixed effects:

Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-2.91440	0.24933	-11.689	< 2e-16	***
logVOT	7.86057	0.53264	14.758	< 2e-16	***
lg(j)	3.67556	0.26220	14.018	< 2e-16	***
h1h2	0.59879	0.07128	8.400	< 2e-16	***
f0	0.59494	0.06655	8.940	< 2e-16	***
logVOT:lg(j)	-4.64492	0.56751	-8.185	2.73e-16	***
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 4.6: Table of the coefficients of the mixed effects logistic regression model that predict the transcribed voiced vs. voiceless stops produced by English children's productions (voiced vs. voiceless) using log VOT, H1-H2 and *f0*: Equation 4.2). Non-high vowel context only.

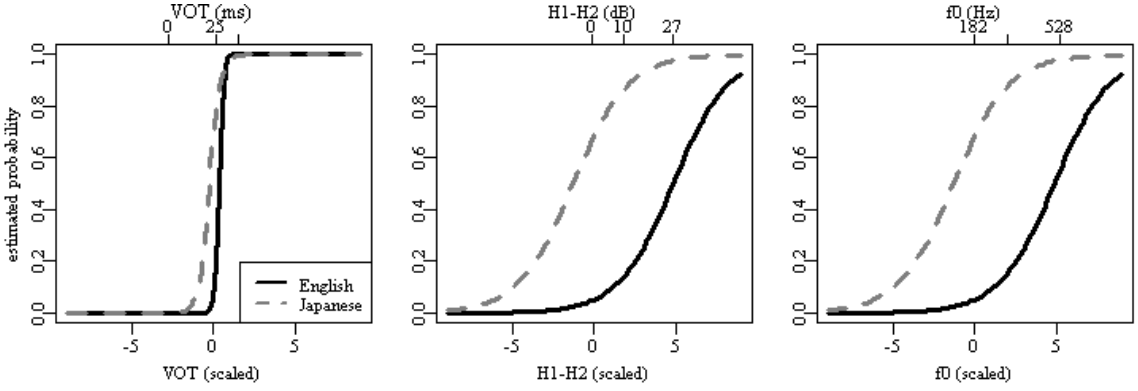

Figure 4.10: The curves of estimated probability with respect to VOT, H1-H2 and *f0* parameters generated from the mixed effects models of logistic regression in English and Japanese children's productions. The estimated probability of '1' indicates the transcribed voiceless, whereas '0' indicates the transcribed voiced. The exact coefficient values are shown in Table 4.6

4.4 Discussion

In this chapter, we examined the acoustic characteristics of voiced vs. voiceless stops in Japanese and English speaking children's productions to explore how they would affect the accuracy judgments in the two languages. The specific question pursued was how the transcribed production accuracy would be determined in Japanese children's stop productions with ambiguous VOT values. In Japanese, observation of the large overlap between VOT values for the voiceless variant of the voiced stop (short lag VOT values) and VOT values for the voiceless stop (intermediate lag VOT values) led us to hypothesize that there might be other acoustic cues which the transcriber used in judging whether children's productions were on target for the voicing feature. To explore this question, we described the relationship between the acoustic properties of target voiced and voiceless stops and the production accuracy of the voicing categories based on the transcriptions, and made a crosslinguistic comparison between English and Japanese.

We considered the acoustic dimensions of VOT, H1-H2 and *f0*. In English, it was primarily VOT that distinguished voiceless stops from voiced stops, with adults using distinct VOT ranges for the two categories (long lag VOT for voiceless stops and short lag or lead VOT for voiced stops). Although there were significant differences between voiced and voiceless stops along the H1-H2 and *f0* dimensions as well, the much better separation along the VOT dimension made these other potential cues comparatively redundant. English-speaking children also produced these stop voicing categories in an adult-like manner, such that voiceless stops were mostly at the long lag end of the VOT scale and voiced stops at the short lag end of the positive VOT region on the scale. In the same vein, the mixed effects logistic regression model showed that the native English-speaking transcriber largely relied on the VOT

properties of the transcribed tokens in making her voicing accuracy judgment. Based on the VOT values of children's stops alone, the model could correctly predict how 93% of the tokens would be transcribed.

In contrast, the voiced and voiceless stops in Japanese were not as well separated from each other along the VOT dimension. The greater overlap between the two category dimension left considerably more room for H1-H2 and *f0* to play a role. The VOT realizations for voiced and voiceless stops produced by Japanese-speaking children patterned similarly to those produced by adult speakers of the language, except they lacked any gender effect. We observed language-specific intermediate VOT values for voiceless stops and the short lag variant for voiced stops in children's stop productions, with no difference between boys and girls across languages.

To summarize, we showed that the contrast between voiced and voiceless stops is characterized differently from language to language. Not every language realizes its voicing or aspiration contrast using two of the three distinct VOT ranges that Lisker and Abramson (1964) defined. In particular, Japanese voiceless stops occupy an intermediate range of values between short and long lag VOT. Moreover, not every language successfully differentiates voiced stops from voiceless stops along the VOT dimension alone. Japanese voiceless stops were not robustly distinguished from voiced stops by their somewhat longer VOT values and differentiation was improved if their higher (breathier) H1-H2 and higher *f0* values were also taken into account. English voiceless stops also have breathier H1-H2 values and higher following *f0*, but the much greater differentiating power of the VOT difference makes these other cues completely redundant for adult productions. These language-specific characteristics of voiced vs. voiceless stops were reflected in ways of assessing children's production accuracy; whereas the English-speaking transcriber relied primarily on VOT in making her judgments, assigning the category that was a better fit by the adult VOT norm,

the Japanese-speaking transcriber could not base her judgments just on the VOT since this parameter does not differentiate the categories robustly even in the adult productions. The voiceless and voiced stops produced by Japanese-acquiring children would be judged as correct only when the Japanese children achieve highly adult-like realizations of the redundant parameters such as H1-H2 and *f0*, and not just VOT.

CHAPTER 5

PERCEPTION OF ENGLISH AND JAPANESE STOPS

Based on the overlapping adult VOT ranges and the late age of mastering voiced stops in Japanese in Chapter 2, we speculated that there must be other acoustic dimensions that complement VOT in characterizing the Japanese voicing contrast for adults. The more subtle differences in these other dimensions may be more difficult for children to contrast, so that their productions do not sound correct even when they match the adult patterns for VOT. This idea was tested building regression models to estimate how well the VOT parameter predicts the transcribed stop category and by comparing the prediction accuracy for Japanese productions to the prediction accuracy for English productions. Results of these models (reported in Chapter 4) supported the idea. However, it is not clear whether the results can be generalized to the population at large, since the transcriptions for each language reflect a single transcriber's judgment. In this chapter, we explore the same speculation using the statistical models introduced in Chapter 4 but with a modification to the dependent variable to overcome this limitation of the models. A perception experiment was designed to get multiple native speaker adult listeners' voicing judgments of a subset of the adults and children's stop productions to replace the target category on the single native speaker phonetician's transcription as the dependent variable in the adult and child models in Chapter 4.

5.1 Method

5.1.1 Materials

A subset of adults and children's tokens of /d/- and /t/-beginning words in English and Japanese was chosen to be presented as stimuli to 21 English-speaking and 20 Japanese-speaking adult listeners. Only the consonant and vowel portion was excised from the selected words to avoid any perceptual bias from the lexicon. The following vowel context varied. The stimuli include tokens produced by children and adults by which we aim to establish the perception norm. Table 5.1 and Table 5.2 show the distributions of talker's age and vowel context of the stimuli in each language.

In order to balance the number of tokens which are likely to be perceived as /d/ or /t/ in Japanese and English, the single native speaker transcriber's judgments of the token were referred to as a rough estimation of plausible perceptual outcomes. However, the 400 stimuli (350 from children' production and 50 from adults' production) were chosen mainly based on stop VOT, aiming to sample the VOT distribution so as to reflect the whole range of the natural data. Another 10 practice items were chosen, with VOT covering the same range, to be presented at the beginning of the experiment as practice trials. The topmost panels of Figure 5.1 are VOT distributions of all of the English and Japanese stimuli used in the experiments. The distributions of H1-H2 and *f0* are also presented in the figure. As in the whole database, the distributions of H1-H2 and *f0* for the English stimuli appear to be differentiating /d/ from /t/ more effectively than the distributions of these values for the Japanese stimuli. In the analysis, we excluded the responses to the stimuli which put the stops in a high vowel context (/i/ and /u/) because the acoustic parameter of H1-H2 is vulnerable to A1 (the amplitude of the first formant) in this context. .

English		vowel context				
target	talker age	A	I	U	E	O
/d/	2;0-2;11	3	7	7	5	7
	3;0-3;11	10	9	10	9	11
	4;0-4;11	14	16	11	12	14
	5;0-5;11	11	5	3	6	5
	adults	6	5	4	5	5
/t/	2;0-2;11	12	6	2	9	6
	3;0-3;11	11	11	8	9	8
	4;0-4;11	10	14	7	13	10
	5;0-5;11	9	11	6	6	7
	adults	5	6	5	4	5

Table 5.1: The talker age and vowel context distributions of English stimuli sorted by the consonants (/t/ or /d/). 'Target' refers to the target consonant for the adult talkers' stimuli and to the transcribed /t/ or /d/ for the child talkers' stimuli.

Japanese		vowel context				
target	talker age	a	i	u	e	o
/d/	2;0-2;11	9	0	0	10	10
	3;0-3;11	18	0	0	21	18
	4;0-4;11	26	0	0	25	29
	5;0-5;11	20	0	0	14	16
	adults	9	0	0	7	9
/t/	2;0-2;11	11	3	0	6	6
	3;0-3;11	11	3	0	10	11
	4;0-4;11	11	8	0	10	12
	5;0-5;11	8	7	0	11	6
	adults	7	6	0	7	5

Table 5.2: The talker age and vowel context distributions of Japanese stimuli sorted by the consonants (/t/ or /d/). 'Target' refers to the target consonant for the adult talkers' stimuli and to the transcribed /t/ or /d/ for the child talkers' stimuli.

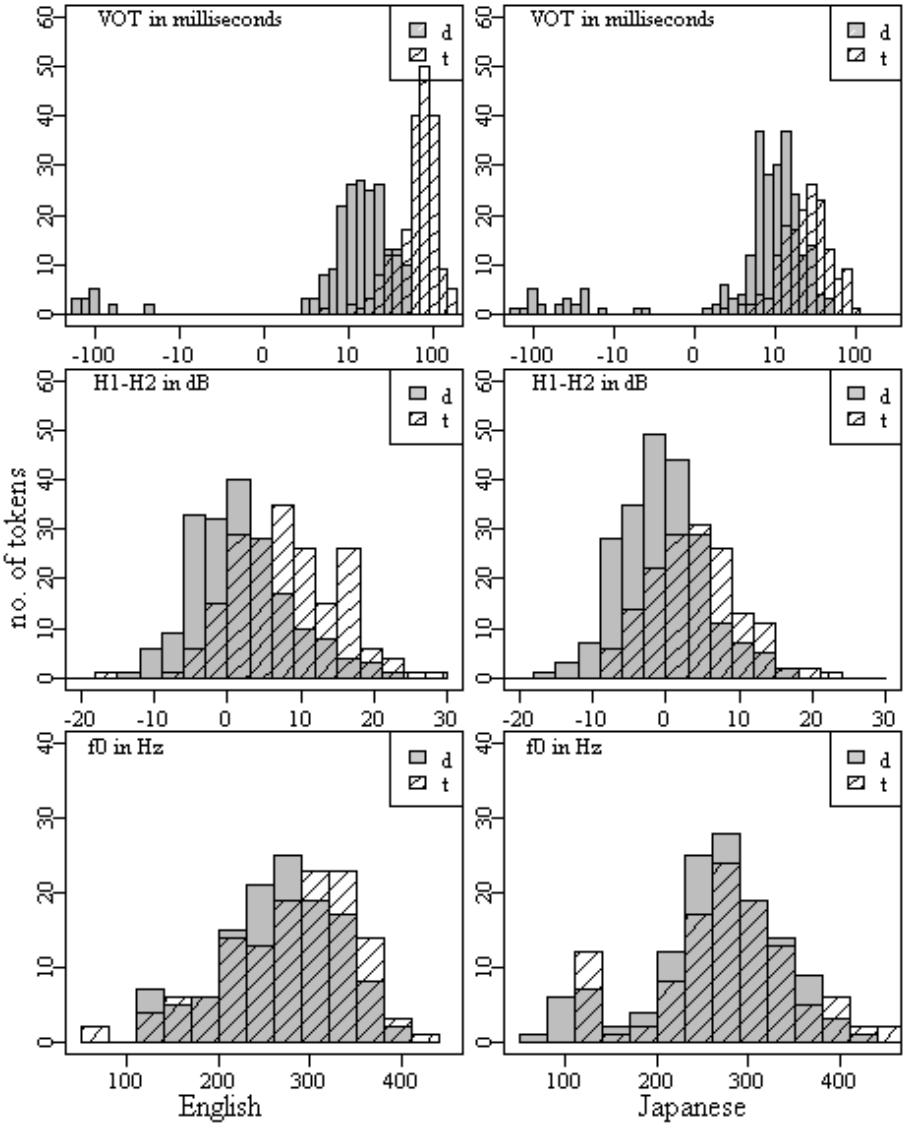

Figure 5.1: Distributions of three acoustic parameters (VOT, H1-H2 and *f0*) in English (left panels) and Japanese (right panels) stimuli.

5.1.2 Subjects and Task

Twenty one college students at the University of Minnesota in Minneapolis and at Daito Bunka University in Tokyo participated in this experiment. They were paid for their participation. The task was to listen to the stimulus on each trial and score it on a visual analogue scale that ranges from 'd' at one end to 't' at the other end. Visual-analogue scaling(VAS) is one type of magnitude scaling method in which participants are asked to assign quantitative numeric values to stimuli in proportion to their magnitude along some perceived scale (Yiu and Ng, 2004). Unlike partition scaling methods such as equal appearing intervals, VAS does not provide listeners with a set of fixed numbers (4- or 7-point scale) but present a continuous visual scale (e.g., 7.5 cm or 10cm long scale) to use in judging the perception of the stimuli. We used a two-pointed arrow for listeners to decide whether the stimulus was closer to /t/ or to /d/.

The procedure was as follows. At the very beginning of the experiment, the double-ended arrow was presented on the computer monitor, as shown in Figure 5.2. Before the real task began, subjects were asked to click on the two end points of the arrow, which are labeled with the two contrastive stop categories ('t' and 'd'), and to click on the mid point of the arrow, which indicates the ambiguous category between the two stop targets. They were informed that they are going to listen to the speech produced by children and adults, and instructed to click at any location along the arrow to estimate the 't'-likeness or the 'd'-likeness of each stimulus. They could select a location at either the 't'-end of the arrow or 'd'-end of the arrow if the stimulus is heard as a clear instance of either of the target end point categories. Or, they could select a location nearer to the midpoint of the arrow if the stimulus is more or less ambiguous. They practiced with the ten practice items before the

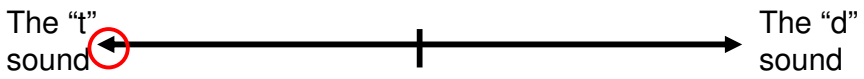

Figure 5.2: Display of double-sided arrows that both English and Japanese listeners used to respond to the stimuli

actual task began. All the instructions were given in writing on the computer screen in the native language of the listener (Appendix B). An optional break was given after every 100 stimuli. All the listeners completed a sequence of two sessions that varied the stimulus language. The first session was for native language stimuli only followed by a session with stimuli from both languages mixed together. We report only the results of the listeners' responses to their native language stimuli.

5.1.3 Statistical analysis

As in Chapter 4, we use mixed effects logistic regression models to describe the relationship between the perceptual responses to the adults' and children's stop productions and the acoustic parameters VOT, H1-H2 and *f0*. As in earlier models, the independent variables are VOT, H1-H2 and *f0* and the linear scale VOT values were transformed into the log scale in the statistical analysis to make the acoustic measure more compatible with the non-linear nature of speech perception.

Unlike the model in Chapter 4, the models in this chapter take as the dependent variable 21 English listeners and 20 Japanese listeners' judgments over the stimuli. This dependent variable is not categorical (i.e., 't' or 'd') but continuous because the listener's judgment is encoded as the pixel number clicked along the arrow in the monitor. The raw pixel numbers that subjects clicked were scaled with a reference to the pixel numbers for the two ends of the arrow (i.e., the calibrated locations of 't' and 'd' as shown in Figure 5.2). To identify the exact pixel number of each end, the pixel numbers that the listeners picked for 't' or 'd' at this calibration stage were examined. The medians of each distribution were taken as the location for 't' and 'd'. The pixel numbers exceeding these medians (i.e., values greater than the median pixel number for 't' and values less than the median pixel number for 'd') were replaced with the medians. Then, the raw pixel numbers were scaled with reference to these two end values, so that '0' reflects the most 'd'-like judgment perception and '1' reflects the most 't'-like judgment.

We separated the responses to the adult talker stimuli from those to the child talker stimuli and built two models in order to first establish the perceptual norms

based on adult productions. Equation 5.1[1] and Equation 5.2[2] show the two models. The difference between the two equations is that the model of responses to adult talker stimuli (Equation 5.1) considers the talker's gender (female vs. male) as an interacting factor with the language (English vs. Japanese) with respect to each acoustic parameter, whereas the model of responses to the child talker stimuli (Equation 5.2) considers the language as the only interacting factor with respect to the acoustic parameter. As in Chapter 4, the three acoustic parameters were standardized using z-score transformation in order to remove the magnitude differences among measurement units (millisecond, decibel and hertz). The z-score normalization was done across both languages. The 'listener' factor was taken as a random effect allowing different variance components to vary according to VOT (log transformed), H1-H2 and *f0* by a listener.

(5.1)

$$log(\frac{\text{`}t' - likeness}{1 - (\text{`}t' - likeness)}) = \beta_1\mathbf{logVOT} + \beta_2\mathbf{logVOT} : \mathbf{Lg.} + \beta_3\mathbf{logVOT} : \mathbf{tk.Gen.}$$
$$+\beta_4\mathbf{logVOT} : \mathbf{Lg.} : \mathbf{tk.Gen.} + \beta_5\mathbf{f0} + \beta_6\mathbf{f0} : \mathbf{Lg.} + \beta_7\mathbf{f0} : \mathbf{tk.Gen.}$$
$$+\beta_8\mathbf{f0} : \mathbf{Lg.} : \mathbf{tk.Gen.} + \beta_9\mathbf{H1H2} + \beta_{10}\mathbf{H1H2} : \mathbf{Lg.}$$
$$+\beta_{11}\mathbf{H1H2} : \mathbf{tk.Gen.} + \beta_{12}\mathbf{H1H2} : \mathbf{Lg.} : \mathbf{tk.Gen.}$$
$$+\beta_{13}\mathbf{Lg.} + +\beta_{14}\mathbf{tk.Gen.} + \gamma\mathbf{Listener}$$

[1] R.code:`lmer(VAS.cal~logVOT.scale*factor(language)*factor(tGender)+f0.scale*factor(language)*factor(tGender)+h1h2.scale*factor(language)*factor(tGender)+(logVOT.scale+f0.scale+h1h2.scale|Subject),data=dat.stat.adult,family=binomial)`

[2] R.code:`lmer(VAS.cal~logVOT.scale*factor(language)+f0.scale*factor(language)+h1h2.scale*factor(language)+(logVOT.scale+f0.scale+h1h2.scale|Subject),data=dat.stat.kid,family=binomial)`

where **Lg.** and **tk.Gen.** refer to the listeners' languages and the talker's gender of the stimuli.

$$
\begin{aligned}
log(\frac{\text{`}t' - likeness}{1 - (\text{`}t' - likeness)}) = \beta_1\mathbf{logVOT} + \beta_2\mathbf{logVOT} : \mathbf{Lg.} + \beta_3\mathbf{f0} \\
+\beta_4\mathbf{f0} : \mathbf{Lg.} + \beta_5\mathbf{H1H2} + \beta_6\mathbf{H1H2} : \mathbf{Lg.} \\
+\beta_7\mathbf{Lg.} + \gamma\mathbf{Listener}
\end{aligned} \tag{5.2}
$$

where **Lg.** refer to the listeners' language.

This mixed effects logistic regression model formulates how much of an effect each acoustic parameter has in predicting the perceptual responses to the stop productions. The coefficients of each parameter in the formula show the relative sizes of the effects. The larger the coefficient, the greater the effect the parameter has in the perceptual judgment of 't'. In order to state the crosslinguistic differences in the effects of each acoustic parameter in judging the stop voicing contrast, we compare the coefficients of the parameter between English and Japanese. Recall that the model is constructed based on the Japanese listeners' responses to the Japanese stimuli, and the English listeners' responses to the English stimuli.

Based on the findings in Chapter 4 regarding the effects of acoustic parameters in differentiating children's voiceless stops from voiced stops, we expect that the effects of VOT in English would be greater than those in Japanese in the naive listeners' perception, either absolutely (the strong version of the hypothesis) or relative to the effects of the other two "redundant" acoustic parameters (the weak version of the hypothesis). That is, the strong version of the hypothesis will be supported if the mixed effects logistic regression model returns a coefficient for VOT in English that is greater than the coefficient for VOT in Japanese with a significant interaction

between VOT and language. Alternatively, the weak version of the hypothesis will be supported if the model returns coefficients for VOT in English that are equal to the coefficients for VOT in Japanese, but greater coefficients for H1-H2 and *f0* in Japanese than in English with a significant interaction between the two variables (H1-H2 and *f0*) and the language, while the coefficients of VOT in English is greater or equal to the coefficient of VOT in Japanese.

5.2 Results

Table 5.3 shows the coefficients of each acoustic parameter (VOT, H1-H2 and *f0*) in the model for the responses to the adult talkers' tokens. There was a significant main effect of language and a significant language by gender interaction. These effects reflected in Figure 5.3 by the separation between the responses by the Japanese- and English-speaking listeners and between the responses of the Japanese-speaking listeners to the stimuli produced by men versus women. Essentially, given the distribution of VOT, H1-H2 and *f0* values in the stimuli, the English-speaking listeners were more likely to choose a point nearer to the "d" end of the arrow and Japanese-speaking listeners rated male stimuli as more "d'-like, other things being equal.

There was a significant main effect of VOT but no significant language by VOT interaction (all three lines in the top panel of Figure 5.3 are equally steep), but there were significant three-way interactions with language and gender for both H1-H2 and *f0*. The Japanese-speaking listeners were much more heavily influenced by these other two parameters in judging the 't'-likeness of the male stimuli.

In summary, the mixed effects logistic regression model suggests that all three independent variables (VOT, H1-H2 and *f0*) influenced English and Japanese listeners' judgments of 't'-likeness when they listened to their native stop voicing categories spoken by adults. While the listeners were affected by VOT to the same degree in

Random effects:	Subject				
Name	Variance	Std.Dev.	Corr		
(Intercept)	4.6789e-10	2.1631e-05			
logVOT	1.2446e-01	3.5279e-01	0.000		
h1h2	1.1581e-02	1.0762e-01	0.000	1.000	
f0	2.7650e-03	5.2584e-02	0.000	-1.000	-1.000
Number of obs:	1059,	groups: Subject,	41		

Fixed effects:					
Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	-1.4143	0.1818	-7.781	7.21e-15 ***	
logVOT	1.5851	0.1999	7.930	2.19e-15 ***	
h1h2	0.8182	0.1358	6.023	1.71e-09 ***	
f0	0.5976	0.1649	3.623	0.000291 ***	
tGdr(m)	5.8960	1.0930	5.394	6.87e-08 ***	
lg(e):tGdr(m)	-4.0110	1.1551	-3.473	0.000516 ***	
tGdr(m):h1h2	0.9739	0.3097	3.144	0.001665 **	
tGdr(m):f0	3.8986	1.1572	3.369	0.000754 ***	
lg(e):tGdr(m):h1h2	-1.0691	0.4605	-2.322	0.020241 *	
lg(e):tGdr(m):f0	-2.5985	1.2620	-2.059	0.039497 *	
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 5.3: Table of the coefficients of the mixed effects logistic regression models for English and Japanese listeners' responses to adult talkers' stimuli. Non-high vowel context only. The table reports only coefficients that are significant at $p<0.05$. 'm' denotes the male talker and 'e' denotes English.

Japanese and English, and for both genders, the Japanese listeners were influenced by H1-H2 and *f0* parameters to a significantly greater degree in listening to the men.

Table 5.4 and Figure 5.4 show the coefficients that were significant at $p< 0.05$. in the mixed effects linear regression model that describes the /t/-likeness ratings of English and Japanese listeners when listening to the child talkers' stops.

There was a significant main effect of VOT, but no significant interaction of VOT with language, showing that English- and Japanese-speaking listeners used

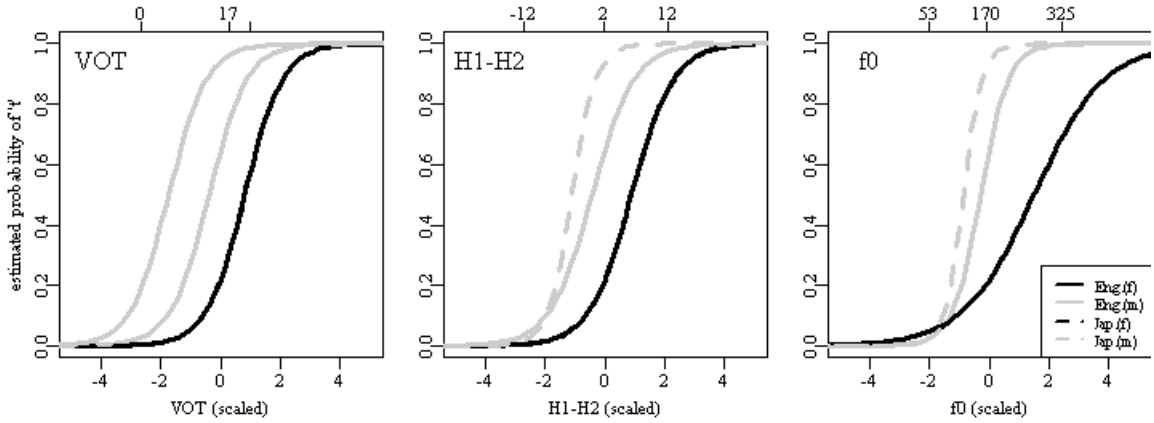

Figure 5.3: The curves of estimated probability in each acoustic parameter from the mixed effects logistic regression model for English and Japanese listeners' responses to the adult talker stimuli (Equation 5.1). The unit normalization was done across two languages. The estimated probability of '1' indicates 't', whereas '0' indicates 'd'. The exact values of the coefficients are shown in Table 5.3.

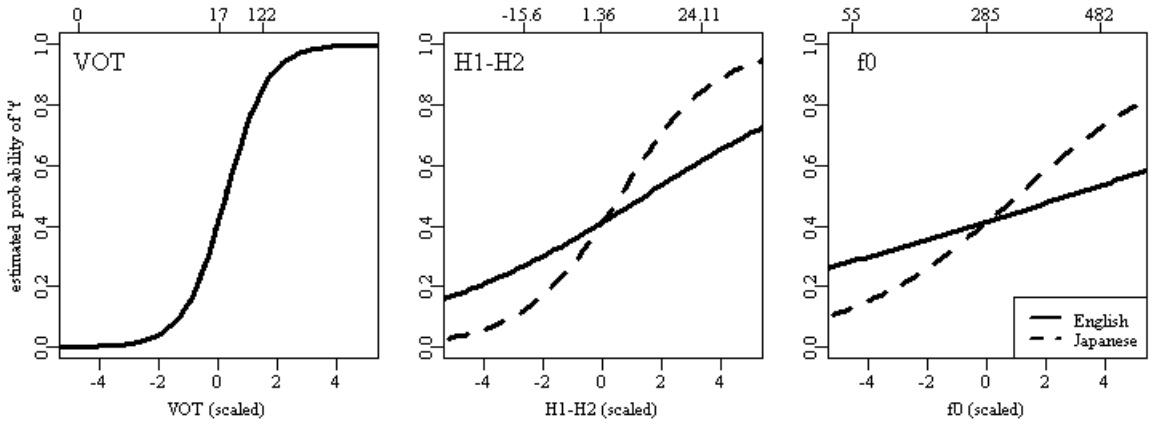

Figure 5.4: The curves of estimated probability in each acoustic parameter from the mixed effects logistic regression model for English and Japanese listeners' responses to the child talker stimuli. The normalization of units was done across two languages. The estimated probability of '1' indicates 't'-likeness judgment, whereas '0' indicates 'd'-likeness judgment. The exact coefficient values are shown at Table 5.4.

Random effects:	Subject				
Name	Variance	Std.Dev.	Corr		
(Intercept)	0.077007	0.277501			
logVOT	0.554666	0.744759	-0.373		
h1h2	0.032787	0.181072	-0.063	0.950	
f0	0.006313	0.079455	0.002	0.927	0.998
Number of obs:	11157,	groups: Subject,	41		

Fixed effects:					
Predictor	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	0.0002874	0.0669461	0.004	0.996575	
logVOT	1.4097611	0.1245281	11.321	< 2e-16 ***	
h1h2	0.6111714	0.0468846	13.036	< 2e-16 ***	
lg(e)	-0.3526486	0.0938064	-3.759	0.000170 ***	
f0	0.3420334	0.0321078	10.653	< 2e-16 ***	
h1h2:lg(e)	-0.3656947	0.0622406	-5.875	4.22e-09 ***	
lg(e):f0	-0.2152828	0.0460156	-4.678	2.89e-06 ***	
Signif. codes: 0	'***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1

Table 5.4: Table of the coefficients estimated from the mixed effects logistic regression models for English and Japanese listeners' responses to child talkers' stimuli. Non-high vowel context only. The table reports only coefficients that are significant at $p<0.05$. The 'e' denotes English.

VOT to the same degree in judging the chidlren's productions. However, there were differences in the role of VOT relative to the other two parameters. Specifically, there were significant interactions between language and H1-H2 and between language and *f0*, such that the coefficients for Japanese-speaking listeners were nearly 3 times the size of the coefficients for the English-speaking listeners. These differences are reflected in the much steeper slopes of the Japanese curves in the two right panels of Figure 5.4. When gauged relative to these language-specific effects, then, the effect of VOT on the Japanese-speaking listeners responses was relatively smaller – 1.40/0.61=

2.3 for VOT relative to H1-H2 and 1.40/0.34 = 4.1 for VOT relative to *f0* as compared to 1.4/(0.61-0.36) = 5.7 and 1.4/(0.34-0.21) = 11.1 for the English-speaking listeners.

Both English and Japanese listeners judged the /t/-likeness of child talkers' stops based on VOT, H1-H2 and *f0* in the non-high vowel context. All three acoustic parameters had a significant influence on the listeners' responses to the stimuli in English and Japanese (significance level: $p<0.05$). The effects of the VOT parameter were not significantly different between English and Japanese, as shown by the identical coefficients.

The effects of H1-H2 were significantly different between English and Japanese ($p<0.05$), in which the coefficient of H1-H2 in Japanese was significantly greater than in English. This suggests that Japanese listeners were influenced by H1-H2 in the judgment of 't'-likeness over the child talker's stop productions more than English listeners were influenced by H1-H2. Likewise, the difference of *f0* coefficients for English and Japanese was significant.

Figure 5.4 shows the probability of 't' as a function of each parameter estimated by the mixed effects logistic regression model of English and Japanese. The left panels plot the distributions of log transformed VOT, H1-H2 and *f0* of English and Japanese stimuli produced by children along the standardized scale. When the coefficients of log VOT, H1-H2 and *f0* in Japanese are overlaid with those in English, the slopes of Japanese curves in H1-H2 and *f0* were steeper than those in English. The steeper slopes of Japanese model indicates the greater effects of H1-H2 and *f0* in Japanese listeners' judgment of 't'-likeness over the Japanese child-talker stimuli than in English listeners' judgment. This suggests that Japanese listeners relied on H1-H2 and *f0* more than English listeners did when they responded to their native stop sounds produced by children.

Figure 5.5 provides the probability curves in three independent variables by each language that are generated by the mixed effects logistic regression model of child-talker stimuli. When the coefficients of three parameters are compared within a language, VOT had the most significant effects in the perceptual judgment of 't'-likeness in both English and Japanese. After VOT, H1-H2 and *f0* were effective in perception in that order.

Compared to English, the relative size of the coefficients among the three parameters in Japanese was smaller. In Japanese, the effect of VOT was 2.30 (=1.4097486/0.6111754) times and 4.12 (=1.4097486/0.3420068) times greater than H1-H2 and *f0*, respectively. In English, the effect of VOT was 5.74 (=1.4097486/0.2454818) times and 11.12 (=1.4097486 /0.1267578) times greater than H1-H2 and *f0*, respectively. The difference in the relative size of the effects among the parameters suggests that Japanese listeners have to rely more on acoustic parameters besides VOT (H1-H2 and *f0*) than as English listeners do in judging 't'-likeness of children's stop productions.

5.3 Discussion

This chapter examined the effects of these acoustic parameters on the responses by English- and Japanese-speaking listeners in a task where listeners were asked to rate the /t/-likeness of CV stimuli when listening to adult or child productions of word-initial stops in their native languages. The response patterns showed that a listener's perception was influenced by various acoustic cues to different degrees, depending on the listener's native language. The results are in agreement with the results of the transcription analyses reported in Chapter 4, in that, while the acoustic parameter of VOT played a large role in predicting the stop voicing category judgments in both English and Japanese, the size of the effect of VOT relative to other parameters (H1-H2 and *f0*) was larger for the English-speaking listeners.

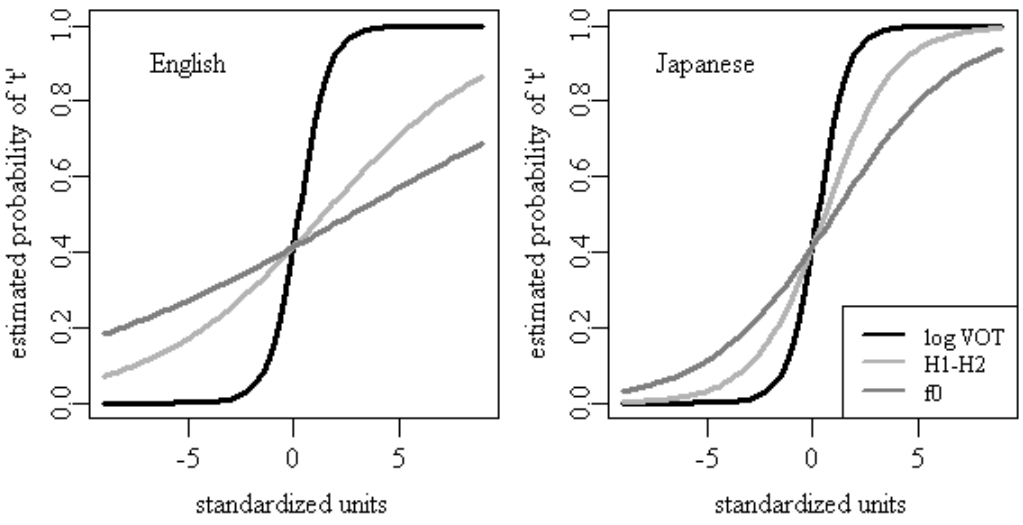

Figure 5.5: The inverse logit curves of the coefficients of each acoustic parameter in English and Japanese mixed effects models of logistic regression. The models were constructed based on the responses to the child talker stimuli only.

These differences are summarized in Figure 5.5, which replots the curves from Figure 5.4 so that the slopes can be compared more easily within each language. Both English- and Japanese-speaking listeners perceived tokens to be closer to the 't' end of the arrow when they had longer VOT, higher H1-H2 and higher *f0*. This is reflected in the positive slopes for all three parameters in both panels of Figure 5.5. VOT was the most influential parameter in describing the listeners' ratings in that the slope for this parameter was steepest. However, H1-H2 and *f0* played a greater role in the responses of the Japanese-speaking listeners. The slopes for these parameters are much steeper in the righthand panel. As a result, the size of the effect of VOT relative to the effects of H1-H2 and *f0* was smaller.

This cross-language difference in the role of the other two parameters suggests an explanation for the differences in the accuracy rates for /d/ productions reported in Chapter 2 and Chapter 4. Specifically, English-speaking children are perceived

as correctly producing /d/ (and not /t/) so long as the VOT is short enough. By contrast, Japanese-speaking children must achieve a less breathy voice and lower *f0* in the following vowel for a token with intermediate VOT to be perceived as /d/.

CHAPTER 6

VOICED STOPS AND NASALS IN GREEK AND JAPANESE

This chapter addresses the puzzling pattern of apparent early mastery of the stop voicing contrast in Greek. Despite the fact that the Greek contrast is between "true" voiceless stops (with short lag VOT) and "true" voiced stops (with lead VOT), the voiced stops are judged as accurate at an earlier age than reported for every other language where stops with voicing lead contrast with short lag stops. As proposed in Chapter 2, this seemingly exceptional pattern of early mastery of Greek voiced stops might be attributed to the prenasalized characteristics of Greek voiced stops. Venting oral air pressure through the velopharyngeal opening is an easy way to achieve the aerodynamic conditions for prevoicing. This would make Greek-acquiring children less subject to the universal constraints that make voiced stops "marked". We test this hypothesis by examining the detailed acoustic characteristics of voiced stops in Greek in order to measure the degree of nasality in the closure.

We assess the prenasalized characteristics of Greek voiced stops by making a within-language comparison to nasal consonants and then by making a crosslinguistic comparison to Japanese voiced stops and nasals. As shown in Chapter 4, the voiced stops in the Japanese adult productions have two variants, as in English, but with the short lag variant being more characteristic of female speakers. In Japanese children's stop productions, the most common acoustic form for voiced stops was the short lag variant, which the native speaker transcriber often judged to be a voicing error. Our

working hypothesis that nasal venting during the closure facilitates the production of prevoiced stops in Greek suggests the following cross-language differences. We predict that Greek-speaking children produce many more tokens of /d/ with lead VOT values. Moreover, these tokens will be more similar to the children's nasals by some acoustic measure of degree of nasal venting during closure. Also, Greek-speaking adults will produce no tokens (or very few tokens) of /d/ that have short lag VOT, and their tokens with voicing lead will (sometimes) show hallmarks of nasal venting during closure. The high incidence of prevoicing will be related to this option for nasal venting. By contrast, the low frequency of prevoiced tokens in Japanese will be related to Japanese children and adults not using nasal venting in producing prevoicing in the voiced stops.

We establish an acoustic norm for degree of nasality by analyzing the nasal consonants in each language and comparing the distribution of our measure of nasality in nasals with the distribution of this measure in the "true" voiced stops in each language in order to assess the degree of nasality in the voiced stop's voice bar (i.e., the interval of lead VOT). This direct comparison between voiced stops and nasals lets us assesses whether Greek voiced stops acoustically resemble nasals more than Japanese voiced stops do.

6.1 Measuring nasality

Although one might suppose that prenasalized stops would have longer lead duration than purely voiced stops due to the more stable supraglottal pressure relative to subglottal pressure, prior studies demonstrate that prenasalized stops are not distinguished from voiced consonants just by duration in Fijian (Maddieson, 1989) or in Moru (Burton et al., 1992). That is, Burton et al. (1992) demonstrated that the duration of the voicing bar in prenasalized stops in Moru does not distinguish them either

from nasals or from purely voiced stops. Similarly Maddieson (1989) measured the acoustic duration of word-medial prenasalized stops in Fijian and found no durational differences between these prenasalized stops and voiced stops or laterals. Therefore, measuring voicing lead would not help to assess the degree of prenasalization in voiced stops in a language. In the current study, we provide the durations of the voice bar or nasal murmur in voiced stops and nasals in Japanese and Greek, but we do not rely on this property to characterize the prenasalized stops as opposed to the purely voiced stops and the nasals. Instead, in defining the prenasalized characteristics of Greek voiced stops in comparison with Japanese voiced stops, we compare the spectral characteristics of the voice bar in voiced stops and those of the nasal murmur in nasals.

The measure we use follows the lead of Burton et al. (1992), who examined acoustic characteristics of phonemically contrastive voiced stops, prenasalized stops and nasals in Moru, a Central Sudanic language. They found that the two types of stops were distinguished from nasals by a falling-off of energy just before the burst, whereas pre-nasalized stops were distinguished from oral voiced stops by higher energy of the nasal murmur during closure. These patterns are schematized in Figure 6.1, which is based on the Figure 6 of Burton et al. (1992).

We adapt the method of amplitude analysis in Burton et al. (1992) so that the nasality in Greek or Japanese voiced stop prevoicing lead is described in relation to the amplitude level of the nasal murmur in nasal consonants. If Greek voiced stops are prenasalized, the amplitude trajectory of prevoicing lead should begin with an energy as high as that in a nasal followed by an energy drop to values distinctly lower than in the nasal murmur. If the voiced stop is not prenasalized, the amplitude trajectory might be at a sustained level over the prevoicing lead interval, but it should be lower than the amplitude of nasal murmur. Finally, a completely nasalized stop would have

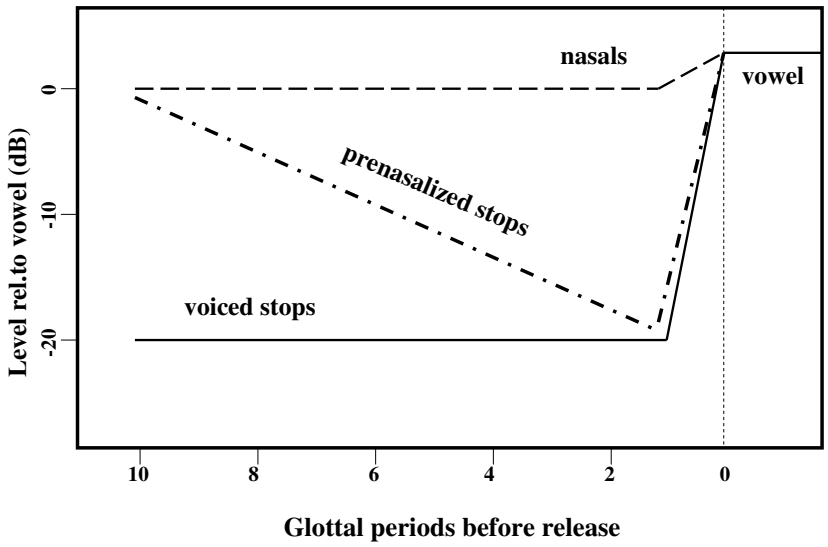

Figure 6.1: The schematized version of phonetic analysis of prenasalized stops in Moru presented in Figure.6 (pp. 137) from Burton et al. (1992). 'The time-course of amplitude changes of the first resonance peak (in dB) prior to the release for nasal consonants, prenasalized stop consonants, and voiced stop consonants. The amplitude values of the glottal periods before the release are normalized relative to the amplitude of the first harmonic of the vowel.'

an amplitude as high as nasal murmur that is sustained during the entire prevoicing lead.

6.2 Method

6.2.1 Subjects

The same child/adult Japanese subjects who contributed to the Database II described in Chapter 2 participated in this data collection as well. Greek participants did not

overlap with those introduced in Chapter 2. The age range of Greek child participants were identical to Japanese children (2;0-6;0). Child and adult subjects were recruited in Tokyo, Japan, and Thesaloniki, Greece. 93 Greek and 80 Japanese children and 6 Greek and 20 Japanese adults participated in the task. Table 6.1 shows the child subjects' age distribution.

6.2.2 Materials and tasks

Voiced stops and nasals were elicited in word-initial position in real words. Target consonants included /b/, /d/, /m/ and /n/ (only /d/ and /n/ were elicited from Japanese adults). Velar stops were excluded from the materials of both languages because of the allophonic variation of Japanese velar nasals as voiced stops (Yamane-Tanaka, 2005).

Children and adults speaking Japanese were given different sets of words because the tasks were different (see Table 6.2,Table 6.3 for Greek speakers's word lists and Table 6.4,Table 6.5 for Japanese speakers' word lists). Although Greek-speaking children and adults also did different tasks to produce words, the target words overlapped between these two age groups because the set of words for children was a subset of the set of words elicited from adults.

The task for the children was originally administered to child participants to assess their consonant articulation ability. Both Greek and Japanese children were asked to name the objects of target words presented in a set of pictures, which also contained other words than the target consonant initial words. When any given picture failed to cue the target word, the researcher directly asked children to repeat after the researcher's voice prompt.

The adults did a slightly different task from the children. Greek target consonants were collected from a word reading task presented on a computer monitor,

language	gender	2;0-3;0	3;0-4;0	4;0-5;0	5;0-6;0	total
Greek	boys	12	12	11	11	46
	girls	10	12	13	12	47
Japanese	boys	10	13	10	8	41
	girls	9	5	15	10	39

Table 6.1: The age distribution of Greek and Japanese child subjects

words	gloss	words	gloss
bala	ball	mixa'ni	motorbikea
du'lapa	wardrobe	ne'ro	water

Table 6.2: The list of words used to elicit Greek word-initial voiced stops and nasals from children.

whereas Japanese target consonants were collected as an appendix to the word repetition task described in Chapter 2.

6.2.3 Acoustic analysis

The duration and amplitude characteristics were measured in the prevoicing lead interval in voiced stops and in the nasal murmur in nasals. The duration was measured based on the interval between two acoustic landmarks of (1) *Prevoicing Beginning* and (2) *Burst Beginning* (= landmark(2) - landmark(1) illustrated in Figure 6.2): *Prevoicing Beginning* was determined based on the first evidence of periodicity in the waveform, and *Burst Beginning* was defined as the first spike of burst energy explosion. As illustrated in Figure 6.2, for nasal consonants, the boundary between the nasal and the following vowel was treated as the analogous location to the burst

words	gloss	words	gloss
b'ala	ball	m'agjes	macho guys
b'eno	get in	m'eno	stay
b'ira	beer	m'inima	message
b'otes	boots	m'onimos	permanent
b'uti	thigh	m'umja	mummy
bal'oni	balloon	magj'es	macho-like actions
berD'evo	confuse	met'a	afterwards
bisk'oto	biscuit	mixan'i	motorcycle
bor'eso	I will be able to	mod'elo	model
buk'ali	bottle	mug'os	dumb
d'ama	queen of hearts	n'ani	sleeping
d'efi	tambourine	n'evro	nerve
d'ini	he/she is dressing	n'ikji	victory
d'opjos	local	n'otos	south
d'uku	to pay in cash	n'umero	number
dal'ika	articulated lorry	nark'ono	drug
deb'uto	first public appearance	ner'o	water
div'ani	divan	nist'azo	drowse
dom'ata	tomato	nom'izo	I think
dul'apa	wardrobe	nuv'ela	short story

Table 6.3: The list of words used to elicit Greek word-initial voiced stops and nasals from adults.

words	gloss	words	gloss
basu	bus	mame	bean
budoo	grape	megane	glasses
		mikaN	orange
deNwa	telephone	naiteiru	crying
		neko	cat
		niNgjoo	doll

Table 6.4: The list of words used to elicit Japanese word-initial voiced stops and nasals from children.

words	gloss	words	gloss
daruma	Dharma doll	namida	tear
deNkji	electricity	nomimono	food
doonatsu	donut	nemuri	sleep
dinaa	diner	numa	swamp
disuku	disk	nihoN	Japan
depaato	department store	nattoo	fermanted beans
daikoN	raddish	nikoniko	smiling
doa	door	nedaN	price
disupuree	display	noriba	platform
doNguri	acorn	nuimono	sewing tools
dekoboko	ragged	nodo	neck
daNgo	rice cake	neko	cat
		nukunuku	snugly
		naSi	pear
		niNgjoo	doll

Table 6.5: The list of words used to elicit Japanese word-initial voiced stops and nasals from adults.

target cons.	2;0-3;0	3;0-4;0	4;0-5;0	5;0-6;0	adults(male)	adults(female)
b	20(18)	24(21)	24(22)	22 (20)	57 (51)	61 (58)
m	12(12)	20(20)	21 (21)	21 (21)	54 (54)	60 (60)
d	15(11)	24 (16)	24(21)	22(14)	52 (50)	58 (56)
n	25(25)	27(26)	22(22)	23 (23)	53(53)	54(54)

Table 6.6: The distribution of Greek voiced stops and nasals collected from child and adult participants that were acoustically analyzed. Numbers in the parenthesis indicate the number of tokens used for the amplitude analysis.

target cons.	2;0-3;0	3;0-4;0	4;0-5;0	5;0-6;0	adults (male)	adults (female)
b	34 (6)	31 (5)	45 (4)	34 (0)	0	0
m	29 (28)	32 (32)	60 (58)	45 (44)	0	0
d	18 (10)	17 (6)	23 (4)	17 (6)	210 (75)	210
n	34 (31)	38 (38)	60 (58)	49 (49)	137 (137)	150

Table 6.7: The distribution of Japanese tokens collected from child and adult participants that were acoustically analyzed. Numbers in the parenthesis indicate the number of tokens used for the amplitude analysis.

of a voiced stop, and the beginning of nasal murmur as the analogous location to the onset of prevoicing in a voiced stop. Therefore, the duration of the voice bar in the voiced stops (i.e., the absolute value of the VOT) corresponds to the duration of nasal murmur in the nasals.

Amplitude values over the voice bar and the nasal murmur were extracted to depict the energy trajectory. Replicating Burton et al. (1992), the amplitude values were measured by taking the first peak amplitude in the FFT spectrum made from a6 ms Hamming window centered at each glottal pulse starting at the burst (*Burst Beginning* as a starting place of window in the burst amplitude) up to *Prevoicing Beginning* as shown in Figure 6.3. The vowel amplitude was measured to normalize the amplitude of the nasal murmur and prevoicing lead. The amplitude of the vowel was obtained by measuring the amplitude of the first harmonic in the spectrum of a 25ms analysis window taken starting at the third pulse after *Burst Beginning*, as suggested by Burton et al. (1992).

The amplitude trajectories of voiced stops and nasals with different numbers of glottal pulses are adjusted to have the same number of data frames (n=20) using a smoothing spline method (Gu, 2002; Davidson, 2006). This method connects and

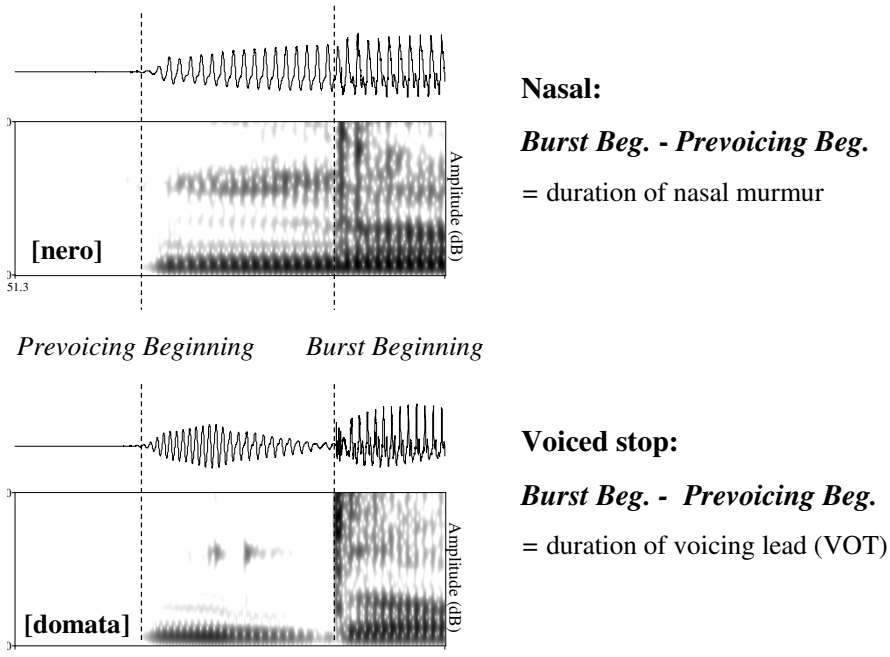

Figure 6.2: Illustration of measuring the duration of prevoicing lead (VOT) in a voiced stop and the duration of nasal murmur in a nasal consonant in Greek.

smooths discrete data points[1]. The 95% confidence intervals of the trajectories were added to the smoothing spline curves to visually inspect the overlap of smoohing spline lines of the categories.

[1]In implementing the analysis, we adapted the R source code from `http://www.u.arizona.edu/~tabaker/ssanova/`. Minor modifications were made in the plotting method.

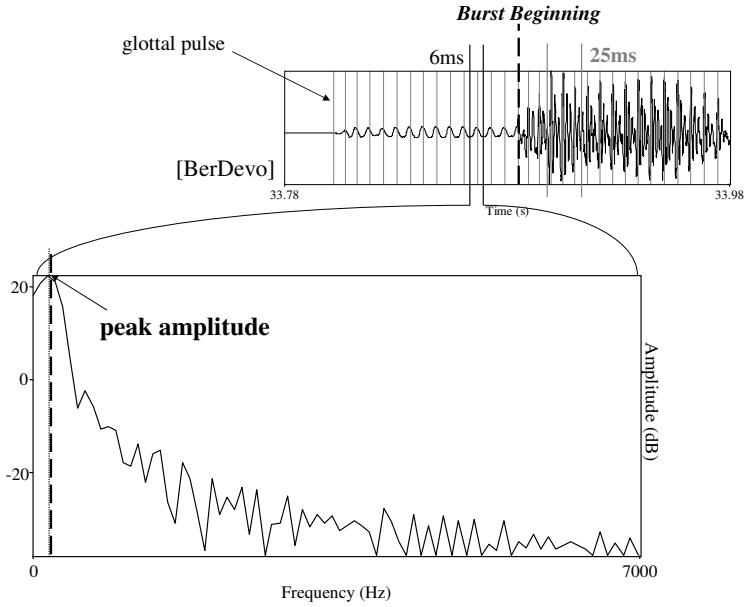

Figure 6.3: Measuring the first peak amplitude of spectrum of 6ms analysis window taken from voicing lead of the voiced stops.

6.3 Results

6.3.1 Greek and Japanese adult productions

Greek voiced stops produced by adults had prevoicing lead in most cases regardless of the gender. Only 5.7% (13/228) of Greek voiced stops were produced without a voicing bar in adult productions. Figure 6.4 shows the histograms of stop voice bars and nasal murmur durations separated by gender. The mean durations of the voice bars in voiced stops (/b/ and /d/) and of the nasal murmur in nasals (/m/, /n/) produced by female speakers were 0.1023 and 0.0877, respectively. The mean

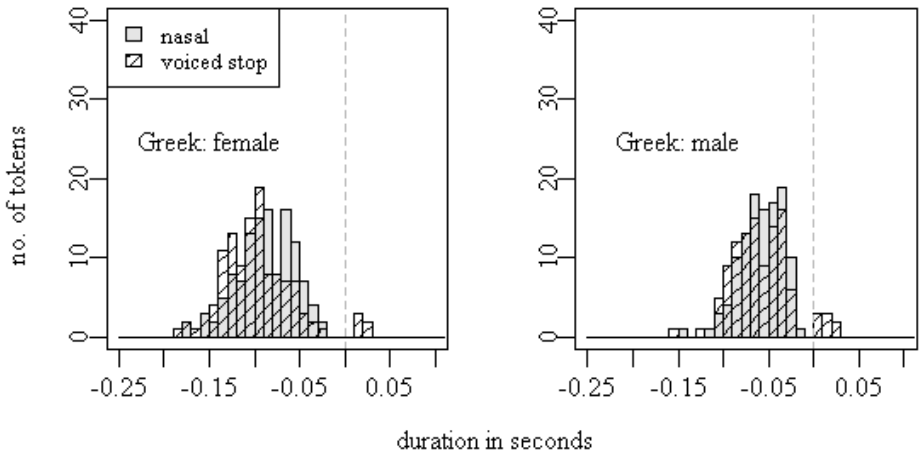

Figure 6.4: Histogram showing the distributions of the duration of the voicing bar in voiced stops and nasal murmur in nasals produced by Greek adults separated by gender.

durations of the voice bar and nasal murmurs produced by male speakers were 0.0577 and 0.0643, respectively.

In contrast, Japanese voiced stops were realized with voiceless variants as well as with prevoiced variants, as discussed in Chapter 2. When separated by gender, 10.4% (22/210) of female's voiced stops and 43.8% (92/210) of male's voiced stops had lead VOT values. Figure 6.5 shows the duration distributions of Japanese voiced stops and nasals separated by gender. The mean durations of the voice bar in /d/ and of the nasal murmur in /n/ were 0.0529 seconds and 0.0433 seconds, respectively. That is, on average the voice bar in Greek voiced stops was longer in duration than in Japanese voiced stops.

Figure 6.6 and Figure 6.7 plot averages and confidence intervals of successive amplitudes measured over a voicing bar in voiced stops and nasal murmur in nasals separated by speaker. The amplitude during a voice bar or a nasal murmur was

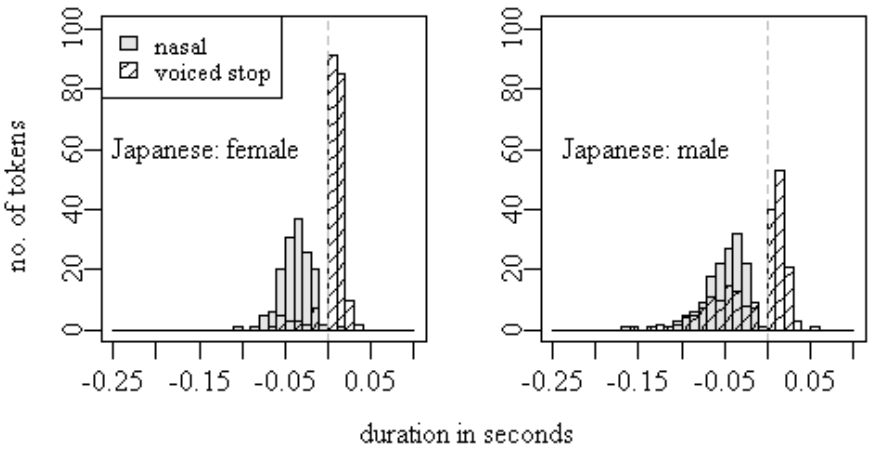

Figure 6.5: Histogram showing the distribution of the duration of the voicing bar in voiced stops and nasal murmur in nasals produced by Japanese adults separated by gender.

normalized with a reference to the amplitude of the following vowel. The abscissa in Figure 6.6 represents the 20 successive time-frames from the beginning of prevoicing lead or nasal murmur (*Prevoicing Beginning* in Figure 6.3) to the location of the burst or the nasal consonant end (*Burst Beginning* in Figure 6.3). The ordinate in the figure represents the normalized sound pressure level of the amplitude with reference to the amplitude of the following vowel. Zero at the ordinate corresponds to an amplitude equal to that in the following vowel.

Each solid trajectory line shows the smoothed mean values of the amplitude measured in the voiced stops and nasals across the glottal pulses prior to the burst. The dotted lines below and above the solid line are 95% confidence interval lines. Each panel compares a set of nasals and voiced stops produced by one of the Greek or Japanese adult speakers.

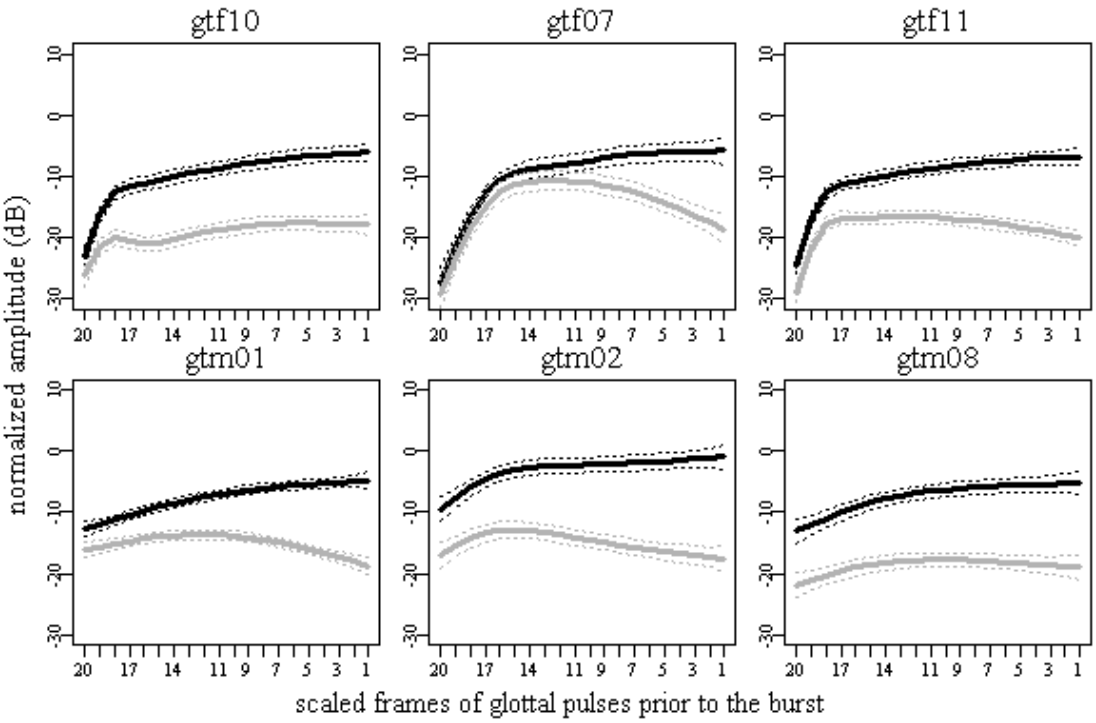

Figure 6.6: The amplitude trajectories of Greek voiced stop lead and nasal murmur elicited at word initial position produced by six Greek adult speakers (voiced stops (/b/, /d/) vs. nasals (/m/, /n/)). The abscissa represents the scaled 20 frames of the glottal pulse arranged in time sequence, and the ordinate represents the sound pressure normalized with reference to the following vowel amplitude.

Across all Greek adult speakers' amplitude curve patterns, nasals were different from voiced stops consistently at the burst region (x=0, rightmost side of the curve) by having a greater amplitude than voiced stops. The nasal murmur amplitude trajectories have a gradual increase over time without abrupt amplitude fluctuations toward the end. The mean amplitude just before the burst referenced with the following vowel amplitude was -4.54 dB (standard deviation= 4.58) for nasals as compared to a mean of only -19.34 dB (standard deviation = 6.43) for voiced stops when averaged across speakers.

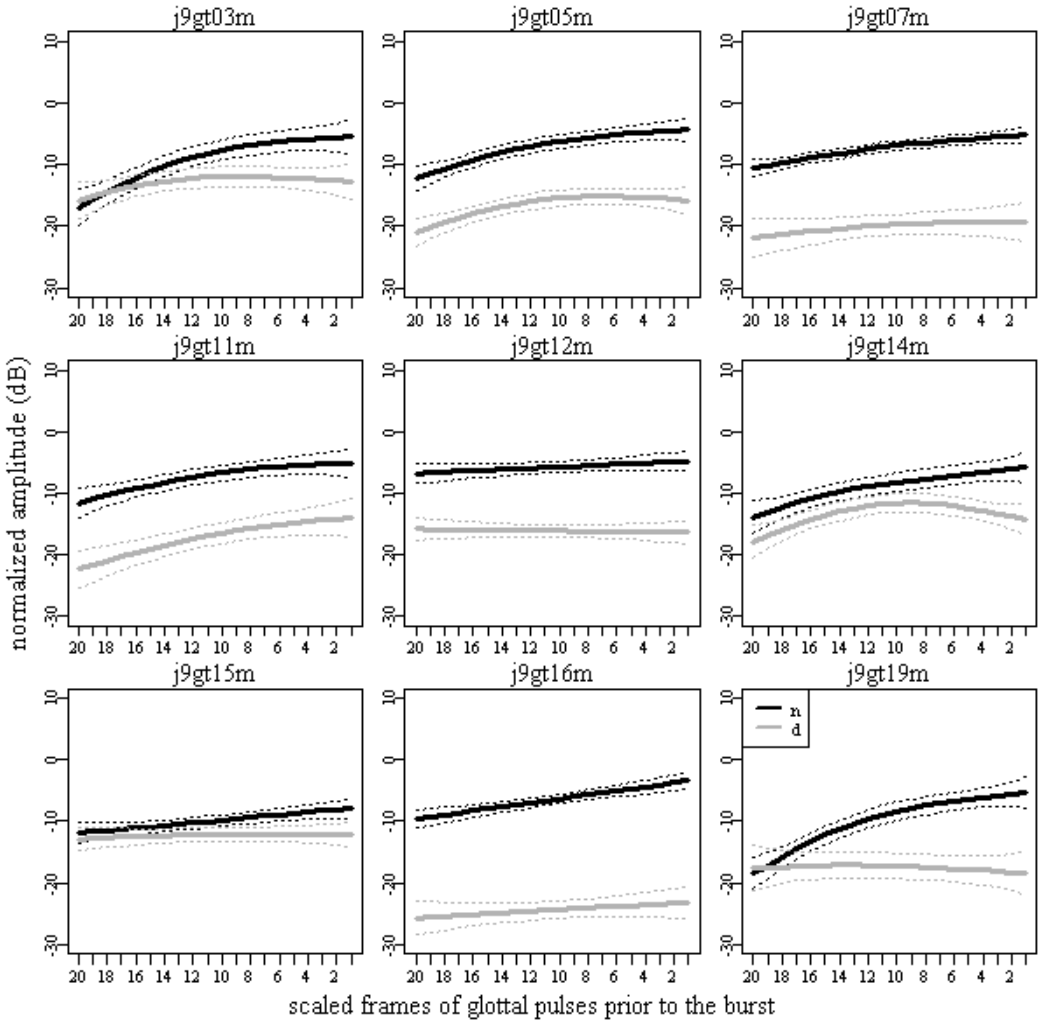

Figure 6.7: The amplitude trajectories of Japanese voiced stop lead in /d/ and nasal murmur in /n/ elicited at word initial position produced by nine Japanese male adult speakers.

In contrast to the consistent shape of the nasal murmur amplitude trajectory, the prevoicing lead of Greek voiced stops showed several different patterns of amplitude trajectories across the six speakers. The difference was primarily at the beginning of the voice bar. Speaker gtf07 in the top center panel of Figure 6.6, for example, began her voiced stops with energy as high as her nasals followed by an energy decrease over time toward the burst. For Speaker gtm08 in the bottom right panel of Figure 6.6, on the other hand, voiced stops began with a lower energy than in his nasals and he maintained this same amplitude difference throughout the voicing bar. The other Greek adult speakers showed amplitude trajectories similar either to Speaker gtf07 or to Speaker gtm08, with minor variation in the relative amplitude at the initial region of the voicing lead. Despite the variation in the amplitude at the initial portion of the voiced stop, the amplitude of the prevoicing lead in Greek voiced stops near the burst was distinctively lower than that of the nasal murmur in the nasal consonants.

Referring to the schematic drawing of amplitude changes over the voicing lead and the nasal murmur in Moru (Figure 6.1), we can say that Greek voiced stops resembled the amplitude characteristics of prenasalized stops in that they have an amplitude decrease approaching the burst. The varying degrees of similarity to the nasal murmur at the initial portion of the prevoicing lead in Greek voiced stops suggests that Greek voiced stops have a prenasalized quality to varying degrees depending on the speaker.

The amplitude trajectories of the voice bars in the Japanese voiced stops and of the nasal murmurs in nasal consonants are shown in Figure 6.7. Only male productions are analyzed since the women had few "true" voiced tokens. Congruent with the Greek nasal amplitude characteristics, the amplitude over the Japanese nasal murmurs gradually increased over the murmur. The normalized nasal murmur amplitude

just before the burst was, on average, -5.7 dB (standard deviation= 3.49), which was consistently higher than the amplitude of the voice bar just before the voiced stop burst (mean= -18.37 dB, standard deviation= 7.24).

The amplitude curves of Japanese voiced stops were different from those of Greek voiced stops for many Greek speakers in that they did not begin with a high amplitude followed by a decrease toward the burst. Instead, the amplitude curves of voice bars in Japanese voiced stops either increased gradually or were at a sustained level parallel to the amplitude trajectories of the nasal murmurs. This relationship between the amplitude trajectories of voiced stops and nasals in Japanese is similar to the relationship between the nasal and the oral voiced stops in Figure 6.1.

6.3.2 Greek and Japanese child productions

In Greek, 84.7% (72/85) of girls' voiced stops and 82.2% (74/90) of boys' voiced stops were produced with prevoicing in word initial position. In each age group, the occurences of prevoiced voiced stops were generally high (85.7%, 79.1%, 87.5% and 81.8% in 2-5 year old groups). The left panels of Figure 6.8 show the distribution of the durations of the voicing interval for voiced stops (i.e., VOT) and of the nasal murmurs in nasals in Greek separately by group. The right panels of Figure 6.8 show the histogram of durations of Japanese voiced stops and nasals. Unlike for Greek, very few voiced stops produced by Japanese children were produced with prevoicing in word-initial position. Only 29.3%(32/109) of girls' voiced stops and 24.5%(27/110) of boys' voiced stops were prevoiced.

The voiced stops and nasals produced by children in both languages mimic the adult patterns in that more Greek voiced stops were prevoiced than Japanese voiced stops. Japanese children rarely produced voice bars in voiced stops, whereas Greek children almost always produced voiced stops with voicing lead.

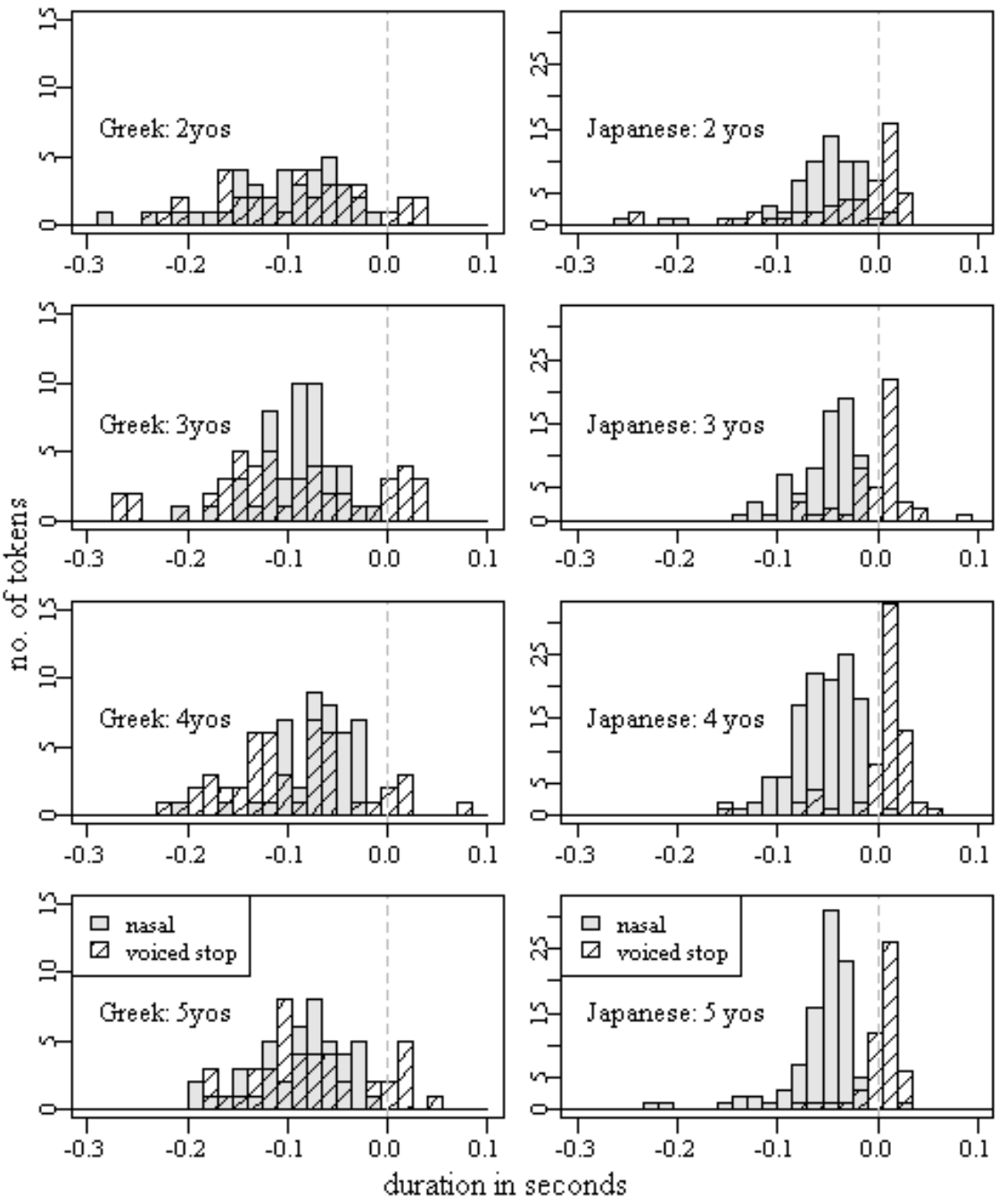

Figure 6.8: Duration of the voice bar and nasal murmur elicited at word initial position by Greek children (left panels) and Japanese children (right panels) separated by the age groups (2;0-6;0).

Figure 6.9 shows the mean amplitude trajectories of the voice bar in voiced stops and of the nasal murmur in nasals produced by Greek children (left panels) and Japanese children (right panels). As in Figure 6.6 and Figure 6.7, the amplitudes are relativized to the amplitude of the following vowels. The solid lines are smoothed mean amplitude of each consonant and the dotted lines are 95% confidence intervals. Since there are fewer child productions, the trajectories are averages over productions by all children in an age group rather than averages by speaker.

Similar to Greek- or Japanese-speaking adult productions of nasals, the amplitude curve of nasal murmur in nasals produced by Greek children lacked an abrupt amplitude excursion during the murmur. This was true for all four age groups. The amplitude immediately before the burst location (x=0) in nasals was higher than the amplitude in voiced stops, although the difference between the two consonant types was smaller than the difference in adult productions. The means of normalized amplitude just before the burst location were -6.64 (standard deviation=3.86) in nasals and -12.55 (standard deviation=7.64) in voiced stops in Greek children's productions.

Greek children's voiced stops were similar to the Greek adult pattern of amplitude trajectories in prevoicing lead in showing an amplitude as high as the nasal murmur amplitude followed by a gradual separation toward the burst. Compared to adult voiced stops, however, the amount by which amplitude dropped approaching the burst was smaller, resulting in a less differentiated amplitude level between voiced stops and nasals at the burst region. This amplitude pattern over time suggests that many of the Greek children's voiced stops have a strong nasal quality throughout the voicing lead interval. The degree of nasality during the duration of voicing lead varied in a way that some voiced stops had a prenasalized property at the initial portion of the voicing lead, and many others had a fully nasal-like quality over the entire

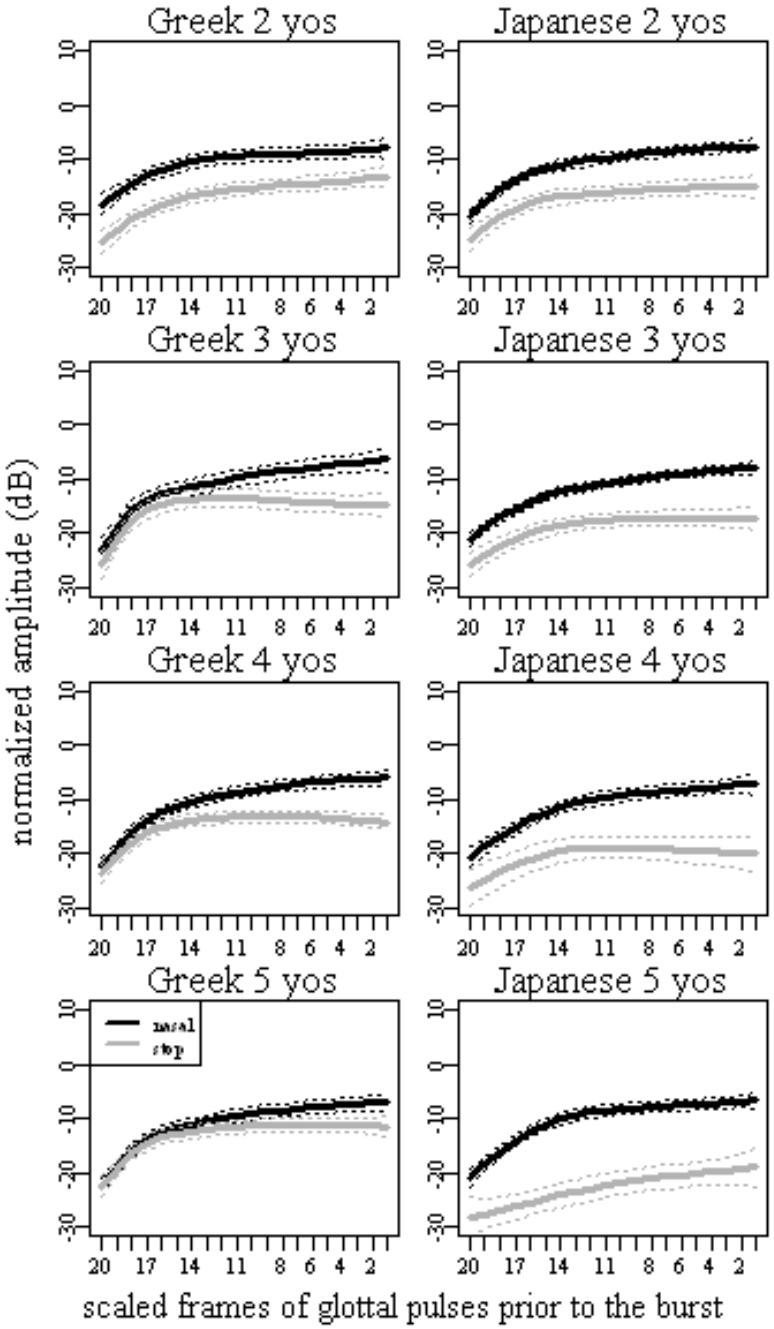

Figure 6.9: The amplitude trajectories of Greek voiced stop lead and nasal murmur elicited at word initial position produced by Greek child speakers (left panels) and Japanese child speakers (right panels). The abscissa represents the scaled 20 frames of the glottal pulse arranged in time sequence, and the ordinate represents the sound pressure normalized with a reference to the following vowel amplitude.

duration of the voicing lead. Therefore, Greek children's voiced stops are characterized either as strongly prenasalized voiced stops or as almost nasal-like voiced stops ("post-stopped nasals").

The right panels in Figure 6.9 shows the amplitude trajectories of Japanese children's nasals and voiced stops. Across all four age groups, the amplitude of nasal murmur gradually increased just as in the nasals produced by Greek and Japanese adults. Immediately before the burst, the amplitude of nasals was consistently higher than that of voiced stops where the amplitude means were -7.2 (standard deviation= 3.37) in nasals and -16.01 (standard deviation= 7.04) in voiced stops.

The difference in the amplitude trajectory pattern between Greek and Japanese children's voiced stops was that the amplitude values of Japanese children's voiced stops were lower than those of their nasals over all frames. The amplitude in the Japanese children's voiced stops remained relatively level over the duration of the prevoicing lead, and this level was lower than the amplitude in the nasals from the beginning. Although the initial amplitude of the trajectories in Japanese children's voiced stops was closer to the amplitude of nasals, the amplitude of voiced stops was clearly lower than that of nasals immediately before the burst. This pattern of amplitude trajectories is similar to the Japanese adults' voiced stops which we characterized as purely voiced stops based on Figure 6.1.

6.4 Discussion

We examined the acoustic characteristics of voiced stops in Greek and Japanese to test the hypothesis that the earlier mastery of Greek voiced stops with adult-like prevoicings might be related to the prenasalized characteristics of Greek voiced stops. In terms of our broader aim, we provided more evidence of fine-grained differences

for purportedly identical categories across languages, which helps us to understand a difference in the mastery patterns between the target languages.

The amplitude trajectory of the prevoicing lead seemed to capture the pre-nasalized quality in voiced stops of Greek fairly accurately. When compared with the amplitude trajectories of the nasal murmur in nasal consonants, Greek voiced stops revealed their prenasalized property by beginning the prevoicing lead with an amplitude as high as in the nasal murmur, followed by an amplitude drop toward the burst. Although the degree of similarity to nasals varied depending on the speaker, the amplitude of Greek voiced stops immediately before the burst was consistently lower than the nasal amplitude across speakers. This prenasalized property was a language-specific characteristic of Greek voiced stops that Japanese voiced stops did not share.

Children acquiring Greek or Japanese also showed a language-specific realization of voiced stops, which resembled the language-specific voiced stop characteristics in the adult productions for the same language. While the prevoiced stops by 3-5 year old Japanese children lacked the prenasalized characteristics just as Japanese adult stops did, Greek children produced prevoiced stops with varying degrees of nasality. In addition to mimicking the language-specific characteristics of having pre-nasalized voiced stops, Greek children showed a high degree of nasality that was extended over the entire duration of the prevoicing lead. Thus, Greek children's voiced stops might be characterized as strongly prenasalized.

The relative frequency of prevoicing in voiced stops between Greek and Japanese must be affected by this crosslinguistic difference in the degree of nasal venting voiced stops. As we hypothesized earlier, the use of the pharyngeal port opening would help children to overcome the difficulty of maintaining a high enough supraglottal pressure throughout the closure, facilitating the production of prevoicing in voiced stops.

The evidence of a prenasalized or nasalized quality captured in the prevoicing lead by Greek children supports the hypothesis that Greek children take advantage of the prenasalized nature of the native language voiced stops and thus achieve an early production proficiency in voiced stops. In contrast, Japanese children's voiced stops lacked prenasality or nasality, and they produce only a small number of voiced stops with prevoicing.

When the acoustic nature of the language-specific voiced stops is taken into consideration, the puzzling early mastery of the Greek voiced stops need not be treated as an exception to developmental universals any more. Consistent with findings from previous chapters, it is necessary to understand the language-specific acoustic details of categories in order to generalize the crosslinguistic patterns in phonological acquisition. In order to understand these details, relevant acoustic aspects other than VOT often need to be considered. Here the amplitude characteristics of prevoicing lead were critical in characterizing prenasality in crosslinguistic categories of voiced stops.

CHAPTER 7

CONCLUSION

This dissertation investigated how a better understanding of language-specific acoustic characteristics of stop phonation-type categories might help to explain cross-language differences in children's production patterns, as evaluated by skilled native speaker transcriber. The conventional wisdom is that whenever a language contrasts a more common or less precisely specified consonant type with a more unusual or more highly specified consonant type, the more common "default" type should be mastered first. In Japanese, for example, "plain" [k] is produced accurately before palatalized [k^{j}] (Nakanishi, Owada, and Fujita, 1972), and in Xhosa, plain velar and alveolar plosives are mastered earlier than and substitute for clicks (Mowrer and Burger, 1991). This conventional wisdom predicts that unusual phonation types such as the tense (or mildly ejective) stops of Korean should be mastered later than the corsslinguistically more common "plain" (or lax) types. It also predicts the cross-linguistically common observation that voiceless unaspirated stop appear in children's productions earlier than voiced or aspirated stops. Voiceless unaspirated stops can produced with many different degrees of glottal opening at any time during and just after the stop release. By contrast, production of aspirated stops requires that the glottis be opened wide at the release of the oral constriction and that the adduction gestures for the following vowel be timed fairly precisely. Also, because initiation and maintenance of voicing requires a pressure drop across the glottis, the production of "true" voiced stops (i.e.,

stops that are specified for voicing lead) requires some extra gesture to vent pressure just enough to achieve the pressure drop, without losing the oral pressure buildup for an audible stop release. For language with "garden-variety" voicing and aspiration contrasts, therefore, the adult VOT norms often are a useful acoustic measure for predicting the relative order in which children master the stop phonation types.

This dissertation examined, three specific cases in which the conventional wisdom fails. The first was the early mastery of the Korean tense stop, which has a high production accuracy rate despite the complex articulatory specifications of posturing the larynx and other articulatory systems for positive vocal cord tension and vocal tract tension. The second was the late mastery of Japanese voiced stops, many of which are transcribed as incorrect even when they are produced as short lag stops that were well within the adult range for the short lag variant of this type. The third was the early mastery of Greek voiced stops, which even very young children produced with highly adult-like lead VOT values. In order to understand these three puzzles, we had to look more carefully at the VOT patterns and we had to look beyond VOT to understand the role of other acoustic parameters.

First, the early mastery of the tense stop in Korean-speaking children's productions was examined in light of a sound change in progress that is reshaping the older distinction among "tense" stops (with non-modal voice after short lag), "plain" stops (with a somewhat longer short lag), and "aspirated" stops (with long lag). In the young Korean adult speakers' stop productions that we examined, it was mainly VOT that separated the tense stop from the "plain" lax stops, which now have long lag VOT values that are more or less completely overlapped with the VOT values for aspirated stops. In differentiating the lax stops from the aspirated stops in the phrase-initial position where we elicited them, *f0* played a critical role in adult productions.

In the Korean adult speakers' stop productions, it was mainly the VOT parameter that separated the tense stop from the lax or the aspirated stops, reflecting the recent sound change that has resulted in a longer lag VOT for the lax stops. The lax stops did not overlap with the tense stops but overlapped with the aspirated stops along the VOT parameter. In differentiating the lax stop from the aspirated stops, *f0* played a critical role in adult productions.

This new production pattern affects the native transcribers' judgments of the accuracy of the tense versus lax categories as they are produced by children. A stop with the "default" short lag VOT was transcribed as correct for the tense stop more readily than for the lax stop. The transcribed categories for children's lax and aspirated stops tended to be "tense" for short lag values, and otherwise largely determined by the *f0* value differentiating the productions that were transcribed as lax stops from those that were transcribed as aspirated. The finding that the VOT determined the transcriber's accuracy judgment of children's tense stop productions helps explain why the tense stop has such a high production accuracy, contrary to the conventional wisdom.

The Japanese puzzle, like the Korean one, involves a change in progress. Japanese voiced stops in many young adult women's productions do not have voicing lead. This short lag variant makes it seem as if the Japanese voicing contrast is becoming a contrast between short lag and longer lag. However many of the children's productions of the voiced stops were transcribed as voiceless, even though the VOT values were well within the adult norms for the new short lag variant. Unlike the English contrast between voiced and voiceless stops, however, the Japanese contrast is not well defined by reference to only VOT. That is, the voiced and voiceless stops are not well differentiated along the VOT continuum. Instead, Japanese voiceless stops are realized with lag VOT values that are intermediate between the short lag VOT

and the long lag VOT categories of Cantonese, English and other language with aspiration contrasts investigated by Lisker and Abramson (1964). While even the short lag VOT tokens of voiceless stops are transcribed as voiceless, the short-lag voiced stops were transcribed as incorrect by the native speaker. To resolve this puzzle, we examined the role of two other acoustic parameters (H1-H2 and *f0*) in a regression model that compared native speaker transcriptions and perceptual responses to Japanese- and English-speaking children's productions. Although VOT was a very important predictor for both the trained and the naive listeners for both languages, H1-H2 and *f0* played a larger role in determining the transcriptions and the naive listeners' responses. These findings suggest that Japanese voiced and voiceless stops are defined differently from English voiced and voiceless stops in acoustics spaces. The language-specific characteristics of stop voicing contrasts affect native listeners' evaluation of children's production accuracy.

Third, we explored the exceptionally early mastery of the Greek voiced stops in children's productions. Greek voiced stops are characterized by their lead VOT values which contrast with the short lag VOT values for the voiceless stops. Although this VOT property of Greek voiced stops would predict the mastery of voiced stops at a relatively later age, following the universal trend of late mastery of 'true' voiced stops in Spanish (Macken and Barton, 1980b), French (Allen, 1985), and Thai (Gandour et al., 1986), the transcribed accuracy of Greek voiced stops was relatively high even in two year olds' productions, most of which had adult-like prevoicing (lead VOT). We investigated how this high production accuracy is related to the language specific characteristics of the Greek voiced stop, which is known to have a prenasalized variant. We hypothesized that this prenasalization would facilitate the production of

prevoicing in the voiced stops, since an open naso-pharyngeal channel would help children meet the aerodynamic conditions for initiating and maintaining voicing during closure.

By comparing the spectral characteristics of the voicing lead in voiced stops and of the nasal murmur in Greek nasals, we could capture the fact that the amplitude characteristics of the voice bar at the beginning of voiced stops are highly similar to those of nasals in Greek. The Greek-acquiring children were mimicking this partially nasalized property of voiced stops, and further extended the nasal quality to the entire duration of the voicing lead. We inferred that Greek children overcome the difficulty of prevoicing through the use of nasality.

A consistent finding across the latter two cases was that even the cross-linguistically common phonological stop voicing contrast is realized in the acoustic space in a language-specific way. Although Japanese voiced stops are becoming like English voiced stops in allowing a short lag variant. Japanese voiceless stops are different from English voiceless stops with respect to the range of VOT values. The English voiceless stops in our study were consistently realized in the long lag VOT range whereas Japanese voiceless stops were consistently in an intermediate lag VOT range. The intermediate lag VOT might be understood as a fourth type of VOT setting - a "partially" aspirated stop - which was previously noted in Canadian French (Caramazza et al., 1973; Caramazza and Yeni-Komshian, 1974), in Hebrew (Raphael et al., 1995) and in Korean as described in older studies predating the sound change (e.g., Lisker and Abramson 1964; Han and Weitzman 1970)

The Greek voicing contrast also taps language-specific categories. Greek voiced stops are different from Japanese and English voiced stops in having a partially nasalized variant rather than a short-lag variant. VOT alone was not adequate for capturing the difference between Greek and Japanese voiced stops.

These subtle acoustic differences for purportedly identical categories have implications for acquisition. Children who learn voiced stops or voiceless stops in different languages have to learn the specific acoustic parametric values appropriate for the ambient language. In keeping with the language-specific acoustic characteristics of the native stop category, children can be delayed or facilitated in reaching production proficiency for the category. Some seemingly exceptional patterns of earlier-than-usual or later-than-usual mastery of some speech categories turn out to not deviate from our expectations when we make an appropriately fine-grained acoustic characterization of the category. The Korean case also reinforces this point. If the lax stop were the "plain" type, as it may have been when Lisker and Abramson (1964) described its intermediate lag distribution, we would expect it to be mastered earlier than the aspirated or the tense categories. However, the recent sound change that shifted the lax stop more into the long lag range has changed the terms of the contrast, so that now the tense stop is more like a default type.

More generally, our investigation of these three seemingly exceptional cases suggest that the acoustic parameters used in describing children's stop productions need to be tailored to the language-specific categories. VOT is not a sufficient acoustic dimension to characterize the transcriber's judgments of the voiced and voiceless stops in Japanese, the voiced stops in Greek, and the lax and aspirated stops in Korean. For these contrasts, the consideration of various other acoustic parametric spaces was necessary to understand how children realize the categories in contrast.

To conclude, there exist various language-specific patterns in the acoustic realization of the stop phonation-type contrasts. While the three-way divisions of the VOT dimension into a voiced type (with lead values), a "plain" voiceless type (with short lag values), and an aspirated type (with long lag values) can sufficiently characterize the phonation-type contrast(s) in some languages, VOT is not always sufficient

for discriminating among stop phonation types in all languages. The implication of this fact is that predictions about children's mastery of a stop phonation-type contrast need to be based on an understanding of the language-specific acoustic nature of the categories in contrast. The seemingly exceptional patterns seen in Korean, Japanese, and Greek might be explained by knowledge of language-specific acoustic details, and further exceptions may also be explained in the same way. Thus, attention to the specifics can allow us to capture more truly universal patterns in the mastery of stop phonation-type categories.

APPENDIX A

PRODUCTION WORD LISTS: DATABASE I, DATABASE II

The lists of words that were used to elicit the word-initial target consonants are presented with WorldBet transcriptions (Hieronymus, 1994).

WorldBet	gloss	target	WorldBet	gloss	target
ta:i22pei35	thigh	t	ka:tc3tsa:tc35	cockroaches	k
ta:n35	eggs	t	ka:u33tsi:n35	scissors	k
taxu35	peas	t	kaxu35	dog	k
tE:55ti:21	daddy	t	kei55ha:i22jaxn21	robot	k
tE:N55	nails	t	kE:N35	neck	k
tekc5si:35	taxi	t	ki:m33	sword	k
ti:m35saxm55	dim sum	t	keN35tsha:tc3	police	k
ti:n33wa:35	telephone	t	ki:pc35	clip	k
ti:pc35	plates	t	ki:tc3tha:55	guitar	k
t>:i35	bag	t	k>:21k>:55	older brother	k
tou55	knife	t	k>:N33khaxm21	piano	k
toN33	cold	t	ku:35	drum	k
tha:i55	ties	th	ku:i33	tired	k
tha:i33j8:N21	sun	th	ku:n33thaxu35	canned food	k
thaxu21	head	th	koN55zaxi35	dolls	k
thE:kc3	to kick	th	kha:55thoN5phi:n35	cartoon	kh
thE:N55	listening	th	kha:tc5	cards	kh
thi:n55	sky	th	khaxtc5	coughing	kh
theN21	stop	th	khE:35tsaxpc5	ketchup	kh
thi:pc3tsi:35	stickers	th	khei23	standing up	kh
thi:u33seN35	skipping	th	khi:m35	pilar	kh
th>:55ha:i35	flip-flops	th	kheN55kaxi35	chatting	kh
th>:i35	table	th	khi:u55tha:i33l>:N	Q-Taro	kh
th>:N35	candies	th	khi:u21	bridge	kh
			kh>:N33ji:23	to protest	kh
			khu:55Na:21	orthodontics	kh
			khu:i35wa:35	drawing	kh
			khokc5khei21pE:N35	cookies	kh
			khoN21	poor	kh
			khu:tc3wu:21	brackets	kh

Table A.1: Cantonese wordlist: Database I.

WorldBet	gloss	target	WorldBet	gloss	target
dIS	dish	d	tuT	tooth	t
duw	dew	d	tod	toad	t
dal	dolly	d	tost	toast	t
dir	deer	d	tal	tall	t
dud	dude	d	tel	tail	t
de	day	d	tas	tossing	t
dIn&9	dinner	d	tub&	tuba	t
do9	door	d	tul	tool	t
dIg	dig	d	tiT	teeth	t
don& t	donut	d	tap	top	t
dag	dog	d	tak	talking	t
dun	dune	d	tIkl	tickle	t
dEntIst	dentist	d	tep	tape	t
dakt&9	doctor	d	tebl	table	t
do	dough	d	tuw	two	t
dendZ&9	danger	d	titS&9	teacher	t
gIft	gift	g	tow	toe	t
goldfIS	goldfish	g	kIk	kick	k
gaN	gong	g	kedZ	cage	k
gufi	goofy	g	kalIN	calling	k
gIgl	giggle	g	kek	cake	k
got	goat	g	kafi	coffee	k
get	gate	g	kot	coat	k
gItIN	getting	g	ki	key	k
ga9dIn	garden	g	kEtS&p	ketchup	k
gost	ghost	g	kon	cone	k
gis	geese	g	kug&9	cougar	k
gabl	gobble	g	kUkIN	cooking	k
gUdiZ	goodies	g	kUki	cookie	k
gem	game	g	ka9	car	k
gus	goose	g	ko9n	corn	k
			kItIn	kitten	k

Table A.2: English wordlist: Database I.

WorldBet	gloss	target	WorldBet	gloss	target
dˆv	dove	d	khˆ.l&r	color	kh
dAl.fIxn	dolphin	d	khˆ.tIxN	cutting	kh
dAn.ki	donkey	d	khA9	car	kh
dE.z&rt	desert	d	khE.tS&p	ketchup	kh
dEk	deck	d	khek	cake	kh
den.dZ&r	danger	d	khev	cave	kh
dI.gIxN	digging	d	khi	key	kh
di9	deer	d	khI.kIxN	kicking	kh
dItS	ditch	d	khI.tSIxn	kitchen	kh
do.n&t	donut	d	kho.ko	cocoa	kh
do9	door	d	khon	cone	kh
domz	domes	d	khot	coat	kh
dud	dude	d	khu.g&r	cougar	kh
duk	duke	d	khU.ki	cookie	kh
dul	duel	d	khUk.IxN	cooking	kh
gˆm.d9Aps	gumdrops	g	thˆN	tongue	th
gA9.dIxn	garden	g	thA.ko	taco	th
gAlf	golf	g	thAl	tall	th
gI.g&lz	giggles	g	the.& l	tail	th
gI.vIxN	giving	g	thEnt	tent	th
gIft	gift	g	thest	taste	th
			thI.k&l	tickle	th
			thi.pi	tepee	th
			thi.tS&r	teacher	th
			tho9n	torn	th
			thod	toad	th
			thost	toast	th
			thu.n&	tuna	th
			thub	tube	th
			thuT	tooth	th

Table A.3: English wordlist: Database II.

WorldBet	gloss	target	WorldBet	gloss	target
daikoNq	daikon raddish	d	tako	octopus	t
daNqgo	round rice cake	d	tamago	egg	t
daruma	Dharma doll	d	tanukji	racoon	t
demekjiNq	gold fish	d	tebukuro	gloves	t
deNqkji	light	d	tegami	letter	t
deNqwa	telephone	d	tempura	tempura	t
doa	door	d	tiSSu	tissue	t
doonatsu	doughnut	d	tooFu	tofu	t
doNqguri	acorn	d	tomato	tomato	t
gakkji	musical instrument	g	tora	tiger	t
gakkoo	school	g	kaba	hippo	k
gasu	gas	g	kame	turtle	k
geemu	game	g	karasu	crow	k
geNqkji	vigor	g	keekji	cake	k
geta	clogs	g	kemuri	smoke	k
goma	sesami	g	keSigomu	eraser	k
gomi	trash	g	koara	koala	k
gorira	gorilla	g	kodomo	child	k
gumi	gummy	g	kotatsu	leg warmer table	k
gurasu	glass	g	kuma	bear	k
guruguru	spinning	g	kuri	chestnut	k
gjiNqkoo	bank	gj	kuruma	car	k
gjiragjira	glaring	gj	kjimono	kimono	kj
gjitaa	guitar	gj	kjippu	ticket	kj
gjaagjaa	screaming	gj	kjitsune	fox	kj
gjaku	reverse	gj	kjabetsu	cabbage	kj
gjorogjoro	goggling	gj	kjappu	cap	kj
gjoodza	dumpling	gj	kjarameru	caramel	kj
gjuuniku	beef	gj	kjookai	church	kj
gjuunjuu	milk	gj	kjorokjoro	look around restlessly	kj
gjuugjuu	crowded	gj	kjoosoo	race	kj
			kjuuri	cucumber	kj
			kjuukjuuSa	ambulance	kj
			kjuu	nine	kj

Table A.4: Japanese wordlist: Database I.

WorldBet	gloss	target	WorldBet	gloss	target
daikoNq	daikon raddish	d	tako	octopus	t
daNqgo	round rice cake	d	tamago	egg	t
daruma	Dharma doll	d	taNqpopo	dendalion	t
dekoboko	jagged	d	tebukuro	gloves	t
deNqkji	light	d	tegami	letter	t
depaato	department store	d	tempura	tempura	t
doa	door	d	tiikappu	tea cup	t
doNqguri	acorn	d	tiiSatsu	t-shirt	t
doonatsu	dounut	d	tiSSu	tissue	t
gakkoo	school	g	tomato	tomato	t
garasu	glass	g	tooFu	tofu	t
gasu	gas	g	tora	tiger	t
geemu	game	g	kaba	hippo	k
geNqkji	vigor	g	kame	turtle	k
geta	clogs	g	karasu	crow	k
goma	sesami	g	keekji	cake	k
gomi	trash	g	kemuri	smoke	k
gorira	gorrila	g	keSigomu	eraser	k
gumi	gummy	g	koara	koala	k
guruguru	spinning	g	kodomo	child	k
guu	guu	g	koppu	cup	k
gjiNqkoo	bank	gj	kuma	bear	k
gjitaa	guitar	gj	kuri	chestnut	k
gjidzagjidza	jaggedly	gj	kuruma	car	k
			kjabetsu	cabbage	kj
			kjappu	cup	kj
			kjarameru	caramel	kj
			kjiiro	yellow color	kj
			kjimono	kimono	kj
			kjiriNq	giraffe	kj
			kjoosoo	race	kj
			kjoodai	church	kj
			kjorokjoro	look around restlessly	kj
			kjuukjuuSa	embulance	kj
			kjuuri	cucumber	kj
			kjuu	nine	kj

Table A.5: Japanese wordlist: Database II.

WorldBet	gloss	target	WorldBet	gloss	target
daruma	Dharma doll	d	namida	namida	n
deNqkji	light	d	nomimono	food	n
doonatsu	donut	d	nemuri	sleeping	n
dinaa	dinner	d	numa	swamp	n
disuku	disk	d	nihoNq	Japan	n
depaato	department store	d	nattoo	fermanted beans	n
daikoNq	daikon raddish	d	nikoniko	nikoniko	n
doa	door	d	nedaNq	price	n
disupuree	display	d	noriba	platform	n
doNqguri	acorn	d	nuimono	sewing kit	n
dekoboko	jagged	d	nodo	neck	n
daNqgo	daNqgo	d	neko	cat	n
			naSi	pear	n
			nukunuku	snugly	n
			niNqgjoo	doll	n

Table A.6: Japanese voiced stop/nasal wordlist (Chapter 6).

WorldBet	gloss	target	WorldBet	gloss	target
tal.phEN.i	snail	t	ka.baN	bag	k
tan.tShu	button	t	kam.dZa	potato	k
taN.gixn	carrot	t	kaN.a.dZi	puppy	k
ta.ram.dZwi	squarrel	t	ka.wi	scissor	k
tiN.gul.diN.gul	rolling about	t	ki.tSha	train	k
tiN.doN.dEN	bell-ringing sound	t	ki.rin	giraffe	k
tuN.gixl.gE	round	t	ku.du	shoe	k
tul	two	t	kuk.dZa	scoop	k
tu.bu	tofu	t	kuk.su	noodle	k
t'al.raN.i	noise maker	t'	k'ak.t'u.gi	raddish kimchi	k'
t'aN.khoN	nut	t'	k'a.ma.gwi	black bird	k'
t'al.gi	strawberry	t'	k'aN.thoN	can	k'
t'a.rix.rixN	tinkling	t'	k'a.man.sEk	black color	k'
t'iN.doN	door-bell sound	t'	k'i.wˆ.jo	sticking in	k'
t'i.t'i.p'aN.p'aN	car-honking sound	t'	k'iN.k'iN	groaning sound	k'
t'uN.bo	plump person	t'	k'ul.k'ˆk	gulping sound	k'
t'uk.t'uk	dripping sound	t'	k'ul.bˆl	honey bee	k'
t'u.k'ˆN	lid	t'	k'ul.k'ul	oink-oink	k'
tha.ol	towel	th	kha.dix	card	kh
tha.i.ˆ	tire	th	khal	knife	kh
tha.dZo	ostriche	th	kha.mE.ra	camera	kh
thak.dZa	table	th	kha.rE	curry	kh
thi.bi	television	th	khi.wi	kiwi	kh
thi.sjˆ.tShix	t-shirt	th	khi.da.ri	tall man	kh
thu.dˆl.thu.dˆl	grumbling	th	khu.khi	cookie	kh
thuk.thuk	beating around	th	khul.khul	snoring	kh
thuN.thuN	stamping	th	khu.sjˆn	cushion	kh

Table A.7: Korean wordlist: Database II.

WorldBet	gloss	target	WorldBet	gloss	target
dadA	chastizement	d	tAksi	classroom	t
dalIKa	truck	d	tAvli	backgammon	t
dAma	hearts (suite)	d	tAvros	bull	t
dedEktiv	detective	d	tElos	end	t
dEfi	tambourine	d	tEras	monster	t
dekOr	dor	d	tEsera	four	t
dInete	give (Pl.)	d	tIGri	tiger	t
dIsko	discotheque	d	tIpota	nothing	t
divAni	sofa	d	tIxos	wall	t
domAta	tomato	d	tOkso	bow	t
dOmino	domino	d	tOnos	tuna	t
dOpCos	local	d	tOpi	ball	t
dUku	bang	d	tUba	somersault	t
dulApa	closet	d	tUrta	cake	t
dUz	shower	d	tUvlo	brick	t
gAfa	gaffe	g	kArta	card	k
gAma	gamut	g	kAsa	big box	k
gAzi	gas	g	kAstro	castle	k
gofrEta	candy bar	g	kOkalo	bone	k
gOlf	golf	g	kOkinos	red	k
gOl	goal	g	kOkoras	rooster	k
gEisa	geisha	gj	kUkla	doll	k
gEla	bouncing	gj	kUn˜a	swing	k
gEmn˜a	reins	gj	kUta	box	k
			kjEdro	center	kj
			kjEfi	fun	kj
			kjErato	horn	kj
			kjAlo	more	kj
			kjIklos	circle	kj
			kjIma	wave	kj
			kjIpos	garden	kj
			kjALa	binoculars	kj
			kjAra	Kiara (cartoon)	kj
			kjOlas	already	kj
			kjOski	kiosk	kj
			kjUpi	big pot	kj

Table A.8: Greek wordlist: Database I.

WorldBet	gloss	target	WorldBet	gloss	target
b'ala	ball	b	m'agjes	macho guys	m
b'eno	get in	b	m'eno	stay	m
b'ira	beer	b	m'inima	message	m
b'otes	boots	b	m'onimos	permanent	m
b'uti	thigh	b	m'umja	mummy	m
bal'oni	balloon	b	magj'es	macho-like actions	m
berD'evo	confuse	b	met'a	afterwards	m
bisk'oto	biscuit	b	mixan'i	motorcycle	m
bor'eso	I will be able to	b	mod'elo	model	m
buk'ali	bottle	b	mug'os	dumb	m
d'ama	queen of hearts	d	n'ani	sleeping	n
d'efi	tambourine	d	n'evro	nerve	n
d'ini	he/she is dressing	d	n'ikji	victory	n
d'opjos	local	d	n'otos	south	n
d'uku	to pay in cash	d	n'umero	number	n
dal'ika	articulated lorry	d	nark'ono	drug	n
deb'uto	first public appearance	d	ner'o	water	n
div'ani	divan	d	nist'azo	drowse	n
dom'ata	tomato	d	nom'izo	I think	n
dul'apa	wardrobe	d	nuv'ela	short story	n

Table A.9: Greek voiced stop/nasal wordlist (Chapter 6)

APPENDIX B

INSTRUCTIONS FOR THE PERCEPTION EXPERIMENT

B.1 Perception task instructions: English

instruction slide 1:

Thank you for participating in this experiment. We are interested in what makes speech sound adult-like or child-like. You will hear a number of consonant-vowel sequences, taken from real words. Specifically, you will listen to consonant-vowel syllables beginning with the "t" sound (like in the words tee, top, and two) or "d" sound (like in the words D, dot, and do). After each stimulus, you will click on a line, where one end of the line represents a perfect "t" sound and the other represents a perfect "d" sound.

Click the mouse to continue the directions.

instruction slide 2:

In the experiment, you will listen to sound files produced by children and adults. They contain words that are supposed to start with the "t" sound, and ones that are supposed to start with the "d" sound. When you hear what you think is a PERFECT "t" sound, click on the line close to where it says "The 't' sound". When you hear what you think is a PERFECT "d" sound, click on the line close to where it says "the 'd' sound."

Click the mouse to continue the directions.

instruction slide 3:

Sometimes, you won't be sure the syllable began with a "t" sound or a "d" sound. In those cases, you should click a place on the line to show whether you thought it sounded more like "t" or more like "d". If the sound wasn't really "t" or "d" but sounded more like "t", then click somewhere on the line closer to the text that says "the 't' sound." If it sounds more like "d," then click closer to the text that says "the 'd' sound."

Click the mouse to continue the directions.

研究に参加していただき、ありがとうございます。私達は、大人らしい音声や子どもらしい音声とはどのようなものであるかということに興味を持っています。皆さんには単語の一部から切り取られたたくさんの子音ー母音の連鎖を聞いていただきます。特に、ティー、トップ、トゥーといったような単語の語頭にみられる“t”で始まる子音ー母音の音節もしくはディー、ドット、ドイツといった単語の語頭に見られる”d”で始まる子音ー母音の音声を聞いていただきます。

毎回、線上をクリックしていただきますが、そこには一端には “t”の音、もう一端は “d”の音を表しています。

インストラクションの続きのためにマウスをクリックしてください。

Figure B.1: **instruction slide 1**: Japanese instruction for the perception experiment.

B.2 Perception task instructions: Japanese

Japanese instructions presented to the Japanese subjects who participated in the perception experiment.

実験では、子どもと大人によって発話された音声ファイルを聞いていただきます。それらは“t”もしくは“d”で始まる単語です。

音声をきいて、パーフェクトな“t”だと思ったら、“The ‘t’ sound”と書いてるところに近いライン上をクリックしてください。もし、パーフェクトな “d”だと思ったら、“The ‘d’ sound”と書いてあるところに近いライン上をクリックしてください。

インストラクションを続けるために、マウスをクリックしてください。

Figure B.2: **instruction slide 2**: Japanese instruction for the perception experiment.

時々、音節が“t”で始まるのか“d”ではじまるのかよく分からないことがあります。その場合には、その音が“t”のような音なのか、“d”のような音なのかをライン上の場所で表していきます。もし、その音がどちらの音かよくわからないけれど、どちらかといったら“t”の音だと思ったら、“The ‘t’ sound”とかいてあるほうに近いほうをクリックしてください。もし、どちらかといったら“d”の音に近い場合には、“The ‘d’ sound”に近いほうをクリックしてください。その近さについてはどのくらいパーフェクトな音に近いかを現すと考えてください。

説明を続けるためにマウスをクリックしてください。

Figure B.3: **instruction slide 3**: Japanese instruction for the perception experiment.

BIBLIOGRAPHY

Allen, G. 1985. How the young French child avoids the pre-voicing problem for word-initial voiced stops. *Journal of Child Language*, 12. 37–46.

Amayreh, Mousa M. and Alice T. Dyson. 1998. The acquisition of Arabic consonants. *Journal of Speech, Language and Hearing Research*, 41. 642–653.

Arvaniti, A. and B.D. Joseph. 1999. Variation in voiced stop prenazalization in Greek. *Ohio State Working Papers in Linguistics*, 52. 203–233.

Benkí, J. 2001. Place of articulation and first formant transition type both affect perception of voicing in English. *Journal of Phonetics*, 29. 1–22.

Brunelle, Marc. 2006. A phonetic study of Eastern Cham register. *Chamic and Beyond. Oceanic Linguistics*.

Burton, M, Blumstein, and K Stevens. 1992. A phonetic analysis of prenazalized stops in Moru. *Journal of Phonetics*, 20. 127–142.

Caramazza, A. and G.H. Yeni-Komshian. 1974. Voice onset time in two French dialects. *Journal of Phonetics*, 2. 239–245.

Caramazza, A., G.H. Yeni-Komshian, E. B. Zurif, and E. Carbone. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of Acoustical Socieity of America*, 54.

Catford, J.C. 1977. *Fundamental problems in phonetics*. Indiana Univeristy Press.

Cho, T., S. Jun, and P. Ladefoged. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30. 193–228.

Cho, T. and P. Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27. 207–229.

Clumeck, H., D. Barton, Macken, and D. Huntington. 1981. The aspiration contrast in Cantonese word-initial stops: data from children and adults. *Journal of Chinese Linguistics*, 9. 210–224.

Cole, J., J.I. Hualde, and K. Iskarous. 1999. Effects of prosodic and segmental context on /g/-lenition in Spanish. *Proceedings of the Fourth International Linguistics and Phonetics Conference*, 575–589.

Dart, S. 1987. An aerodynamic study of Korean stop consonants measurements and modeling. *Journal of Acoustical Society of America*, 81. 138–147.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of Acoustic Society of America*, 120. 407–415.

Davis, K. 1994. Stop voicing in Hindi. *Journal of phonetics*, 22. 177–193.

Davis, K. 1995. Phonetics and phonological contrasts in the acquisition of voicing: Voice Onset Time production in Hindi and English,. *Journal of Child Language*, 22. 275–305.

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

Esposito, Christina. 2004. Santa Ana del Valle Zapotec phonation. *UCLA working papers*, 103. 71–105.

Francis, A.L., V. Ciocca, V.K.M. Wong, and J.K.L. Chan. 2006. Is fundamental frequency a cue to aspiration in initial stops? *Journal of the Acoustical Society of America*, 120. 2884–2896.

Gandour, H. S. H., J., R. Petty, S. Dardarananda, Dechongkit, and S. Mukongoen. 1986. The acquisition of the voicing contrast in Thai: A study of Voice Onset Time in word-initial stop consonants. *Journal of Child Language*, 13. 561–572.

Goldman, R. and M. Fristoe. 2000. *The Goldman Fristoe Test of Articulation-2*. Pearson Assessments.

Gordon, Matthew and Peter Ladefoged. 2002. Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29. 383–406.

Gu, C. 2002. *Smoothing Spline ANOVA Models*. Springer.

Haggard, M., Q. Summerfield, and M. Roberts. 1981. Psychoacoustical and cultural determinants of phoneme boundaries: evidence from trading f0 cues in the voiced-voiceless distinction. *Journal of phonetics*, 9. 49–62.

Han, M. and R. Weitzman. 1970. Acoustic features of Korean /p',t',k'/, /p,t,k/ and /p^h,t^h,k^h/. *Phonetica*, 22. 112–128.

Hanson, H. M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *Journal of acoustical society of America*, 101. 466–481.

Hanson, H. M. and E. S. Chuang. 1999. Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of acoustical society of America*, 106. 1064–1077.

Hardcastle, W. J. 1973. Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics*, 1. 263–272.

Hieronymus, J.L. 1994. *ASCII phonetic symbols for the world's languages: Worldbet*. AT&T Bell Laboratories.

Hillenbrand, J., R.A Cleveland, and R.L. Ericson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37. 769–778.

Hirose, H., C. Lee, and T. Ushijima. 1974. Laryngeal control in Korean stop prodiction. *Journal of Phonetics*, 2. 145–152.

Holmberg, E. B., R. E. Hillman, and J. S. Perkell. 1988. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. *Journal of acoustical society of America*, 84. 511–529.

Hombert, J.-M., J. J. Ohala, and W. G. Ewan. 1979. Phonetic explanations for the development of tones. *Language*, 55. 37 – 58.

Hombert, Jean-Marie. 1977. Consonant types, vowel height and tone in Yoruba. *Studies in African Linguistics*, 120. 173–190.

Homma, Y. 1980. Voice Onset Time in Japanese. *Onseigakkai kaihao*, 163. 7–9.

Honda, K. 1995. Laryngeal and extra-laryngeal mechanisms of f0 control. *Producing Speech: Contemporary issues*.

Huffman, Marie. 1987. Measures in phonation type in hmong. *Journal of Acoustical Society of America*, 81. 495–504.

Iseli, M., Y.L. Shue, and A. Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of Acoustical Society of America*, 121. 2283–2295.

Jakobson, Roman. 1968. *Child Langauge Aphasia and Phonological Universals*. The Hague:Mouton.

Jun, S. 2007. Phonological development of Korean: a case study. *UCLA Working Papers in Phonetics*, 105. 51–65.

Jun, S.-A. 1993. *The phonetics and phonology of Korean prosody*. Ph.D. thesis, Ohio State University.

Jun, S.-A. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology*, 15. 189–226.

Kagaya, R. 1974. A fiberstopic and acoustic study of Korean stops, affricates, and fricatives. *Journal of Phonetics*, 2. 161–180.

Kang, Kyoung-Ho and Susan G. Guion. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America*, 124.

Kang, Kyung-Shim. 1998. On phonetic parameters in the acquisition of Korean obstruents. chicago linguistics society. *Chicago Linguistics Society*, 34. 311–326.

Keating, P. 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11. 277–290.

Kent, R. 1981. Articulatory-acoustic perspectives on speech development. *Language behavior in infancy and early childhood*.

Kewley-Port, D. and M.S. Preston. 1974. Early apical stop production: A voice onset time analysis. *Journal of Phonetics*, 2. 195–210.

Kim, C.W. 1965. On the autonomy of the tensity feature in stop classification (with. special reference to Korean stops). *Word*, 21. 339–359.

Kim, Hyunsoon, K. Honda, and S. Maeda. 2005. Stroboscopic-cine mri study of the phasing between the tongue andthe larynx in the Korean three-way phonation contrast. *Journal of Phonetics*, 33. 1–26.

Kim, M., P. Beddor, and J. Horrocks. 2002. The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, 30. 77–100.

Kim, Midam. 2004. Correlation between VOT and F0 in the perception of Korean stops and affricates. *Proceedings of INTERSPEECH*.

Kim, Minjung. 2008. *The phonetic and phonological development of word-initial Korean obstruents in young Korean children*. Ph.D. thesis, University of Washington.

Kim, M.J. and S. Pae. 2005. The percentage of consonants correct and the ages of consonantal acquisition for 'Korean-test of articulation for children(K-TAC)'. *Journal of Korean Association of Speech Science*, 12. 139–149.

Kim, Y.-T. 1996. Study on articulation accuracy of preschool Korean children through picture consonant articulation test. *Korean Journal of Communication Disorders*, 1. 7–33.

Klatt, D. and L. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male speakers. *Journal of acoustical society of America*, 87. 820–857.

Lewis, Anthony M. 2002. *Contrast maintenance and intervocalic stop lenition in Spanish and Portuguese: When is it alright to lenite?*, 159–171. John Benjamins.

Lisker, L. and A. Abramson. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Words*, 20. 384–442.

Macken, M. A. and D. Barton. 1980a. The acquisition of the voicing contrast in English: A study of Voice Onset Time in word-initial stop consonants. *Journal of Child Language*, 7. 41–74.

Macken, M. A. and D. Barton. 1980b. The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, 7. 433–458.

Maddieson, Ian. 1989. Prenasalized stops and speech timing. *Journal of the International phonetic association*, 19. 57–66.

Maddieson, Ian. 1996. Phonetic universals. *UCLA Working Papers in Phonetics*, 92. 160–178.

Mowrer, D and S Burger. 1991. A comparative analysis of phonological acquisition of consonants in the speech of 2.5-6-year-old Xhosa- and English-speaking children. *Clinical Linguistics and Phonetics*, 5. 139–164.

Nakanishi, Yasuko, Kenjiro Owada, and Noriko Fujita. 1972. Koon kensa to sono kekka ni kansuru kosatsu [results and interpretation of articulation tests for chidlren.].

Ohala, J. 1983. *The origin of sound patterns in vocal tract constraints*, 189–216. Springer-Verlag.

Ohde, R. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of Acoustical Society of America*, 75. 224–230.

Okalidow, Areti, Kakia Petinou, Elena Theodorou, and Eleni Karasimou. 2002. Development of voiced onset time in Greek and Cypriot Greek preschoolers.

Pan, Ho-hsien. 1994. *The voicing contrasts of Taiwanese (Amoy) initial stops: data from adults and children*. Ph.D. thesis, Ohio State University.

Park, H. 2002. The time course of F1 and F2 as a descriptor of phonation types. *Eoneohag*, 33. 87–108.

Pind, J. 1999. The role of F1 in the perception of voice onset time and voice offset time. *Journal of the Acoustical Society of America*, 106. 434–437.

Raphael, Tobin, Faber, Kolier, and Milstein. 1995. Intermediate values of voice onset time. *Producing Speech: Contemporary issues*.

Riney, T., N. Takagi, K. Otaa, and Y. Uchida. 2007. The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, 35. 439–443.

Romero, J. 1995. *Gestural organization in Spanish. An experimental study of spirantization and aspiration*. Ph.D. thesis, University of Connecticut.

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

Shimizu, K. 1989. A cross-language study of voicing contrast. *Studia phonologica*, 23. 1–12.

Silva, D.J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *phonology*, 23. 287–308.

Smit, A.B., L. Hand, J. Freilinger, J. Renthal, and A Bird. 1990. The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55. 779–798.

So, Lydia K. H. and B Dodd. 1995. The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, 22. 473–495.

Stevens, K.N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. *Human Communication: A Unified View*.

Stevens, K.N. 2000. *Acoustic Phonetics*. Cambridge:MIT Press.

Stevens, K.N. and D.H. Klatt. 1974. Role of formant transitions in the voice-voielss distinction for stops. *Journal of the Acoustical Society of America*, 55. 653–659.

Summerfield, Q. 1982. Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America*, 72. 51–61.

Summerfield, Q. and M. Haggard. 1977. On the dissociation of spectral and temporal cues to thevoicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 62. 453–448.

Sundara, M. 2005. Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *Journal of the Acoustical Society of America*, 118.

Takada, M. 2004a. Physiological foundations of differences in articulation of Japanese stops due to generation change. *Journal of the phonetic society of Japan*, 8. 57–66.

Takada, M. 2004b. VOT tendency in the initial voiced alveolar plosive /d/ in Japanese and the speakers' age. *Journal of the phonetic society of Japan*, 8. 57–66.

Templin, M. 1957. *Certain Language Skills in Children*. Univ. Mennesota.

Titze, I. 1989. Physiological and acoustic differences between male and female voices. *Journal of Acoustical Society of America*, 85. 1699–1707.

Titze, I. 1995. Motor and sensory componants of a feedback-control model of fundamental frequency. *Producing Speech: Contemporary issues*.

Wayland, R. and A. Jongman. 2003. Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics*, 31. 181–201.

Westbury, J. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of Acoustical Society of America*, 73. 1322–1336.

Westbury, J. and P. Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics*, 22. 145–166.

Westbury, J and J Parris. 1970. Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. *Journal of Acoustical Society of America*, 47. 625–633.

Whalen, D.H., A. Abramson, L. Lisker, and Mody. 1993. F0 gives voicing information even with unambiguous voice onset times. *Journal of Acoustical Society of America*, 93. 2152–2159.

Whalen, D.H., Andrea G. Levitt, and Luise M. Goldstein. 2007. VOT in the babbling of French- and English-learning infants. *Journal of Phonetics*, 35. 341–352.

Wright, J. D. 2007. *Laryngeal contrast in Seoul Korean*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

Yamane-Tanaka, Noriko. 2005. *The Implicational Distribution of Prenasalized Stops in Japanese*, 123–156. Mouton de Gruyter.

Yiu, Edwin and Chi-yan Ng. 2004. Equal appearing interval and visual analogue scaling of perceptual roughness and breathiness. *Clinical Linguistics and Phonetics*, 3. 211–229.

Zlatin, M. and R. Koeigsknecht. 1975. Development of the voicing contrast: perception of stop consonants. *Journal of Speech and Hearing Reseach*, 18. 541–553.