2017

Interactive Development Of F0 As An Acoustic Cue For Korean Stop Contrast

Gayeon Son
University of Pennsylvania, son.gayeon@gmail.com

Recommended Citation

Son, Gayeon, "Interactive Development Of F0 As An Acoustic Cue For Korean Stop Contrast" (2017). *Publicly Accessible Penn Dissertations*. 2589.
https://repository.upenn.edu/edissertations/2589

Interactive Development Of F0 As An Acoustic Cue For Korean Stop Contrast

Abstract
Korean stop contrasts (lenis, fortis, and aspirated) have been phonetically differentiated by Voice Onset Time (VOT), but with a tonogenetic sound change in progress, the role of fundamental frequency (F0) has been amplified for distinguishing between Korean stop contrasts with the loss of VOT differentiation in young adults' production. The present study explores how F0 is perceptually acquired and how it phonetically operates in toddler speech in relation to Korean stop contrasts according to age. In order to determine the relationship between F0 developmental patterns and age in child stop production, this study uses a quantitative acoustic model to examine the word-initial stop productions of 58 Korean monolingual children aged 20 to 47 months. The production experiment confirmed that VOT is useful for distinguishing fortis stops, but F0 is required for distinguishing between lenis and aspirated stops, and this tendency is significantly related to age. As F0 becomes a determinant acoustic parameter for articulatory distinction, the role of F0 in perceptual distinction was investigated through a perceptual identification test with the F0 continuum. Children were provided with selected lenis-aspirated pairs of images in which they would point to one or the other image in response to given synthetic sounds with different F0 values. This allowed us to observe how phonetic boundaries in the F0 dimension for aspirated stops change with age. A comparative analysis between children's production and perception of F0 indicates that articulatory skills depend on perceived F0 differences depending on the phonemic categories. Additionally, the analysis indicates that once F0 is acquired, VOT differentiation diminishes for distinction between lenis and aspirated stops, and this trade-off between VOT and F0 would occur around the age of 3 years. These findings suggest that phonemic categorization of lenis and aspirated stops should be processed in the F0 dimension and that phonemic processing in perceptual acoustic space is directly linked to phonetic discrimination between the non-fortis stops in production. This study provides experimental evidence for understanding a developmental trajectory of F0 as an acoustic cue for native phonological contrasts.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Linguistics

First Advisor
Mark Liberman

Keywords
Acoustic cue, F0, Language acquisition, Language development, perception and production, stop contrast

Subject Categories
Linguistics

INTERACTIVE DEVELOPMENT OF F0 AS AN ACOUSTIC CUE FOR KOREAN STOP CONTRAST

Gayeon Son

A DISSERTATION
in
Linguistics

Presented to the Faculties of the University of Pennsylvania in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy
2017

Supervisor of Dissertation

____

Mark Liberman
Christopher H. Browne Distinguished Professor of Linguistics

Graduate Group Chairperson

____

Eugene Buckley
Associate Professor of Linguistics

Dissertation Committee
Eugene Buckley, Associate Professor of Linguistics
Daniel Swingley, Professor of Psychology
Jianjing Kuang, Assistant Professor of Linguistics

INTERACTIVE DEVELOPMENT OF F0 AS AN ACOUSTIC CUE FOR KOREAN STOP CONTRAST

Acknowledgments

During my Ph. D. life at Penn Linguistics, I have benefited from being taught by great linguists, and without their help, I could not have completed this dissertation.

First and foremost, I would like to express my heartfelt gratitude to my advisor, Professor Mark Liberman, whose insightful and constructive suggestions never fail to amaze me and to broaden my view in Linguistics. The guidance he gave me in developing research helped me make it through this project. He is the greatest scientist and researcher I have ever known. I could get inspired from the width and depth of his knowledge and even from his enthusiasm for research. Without his guidance, this project could not have been completed.

I also would like to thank to Professor Daniel Swingley, whose timely and detailed feedback and comments were tremendously helpful. This project was the first work to study language-learning toddlers for me, so everything about it was very new to me. When I was planning this project, his guidance made me approach it in the most appropriate way. His insightful and explicit suggestions greatly improved the design of the experiments with toddlers. He was always generous with his time, so I was able to share any findings with him and he did not hesitate to give me valuable comments.

Another thanks must go to Professor Gene Buckley, whose patience and generosity were always encouraging me. His detailed feedback and suggestions were always invaluable. From him, I received intellectual and emotional support in the Linguistics department. And I also would like to thank to Professor Jianjing Kuang. She was always supportive and encouraging, and did not hesitate to share her ideas with me throughout the completion of this work. During the write-up, she always helped me to keep it on the track.

I owe a great debt to my cohorts, colleagues, and friends. They were always supportive and encouraging. With the existence of them, my Ph. D. life would have been full of excitement and inspiration. Last, but of course most, my dearest appreciation goes to my parents, parents-in-law, husband, and my son, Phillip Juwon, for their unconditional love and support.

ABSTRACT

INTERACTIVE DEVELOPMENT OF F0 AS AN ACOUSTIC CUE FOR KOREAN STOP CONTRAST

Gayeon Son

Mark Liberman

Korean stop contrasts (lenis, fortis, and aspirated) have been phonetically differentiated by Voice Onset Time (VOT), but with a tonogenetic sound change in progress, the role of fundamental frequency (F0) has been amplified for distinguishing between Korean stop contrasts with the loss of VOT differentiation in young adults' production. The present study explores how F0 is perceptually acquired and how it phonetically operates in toddler speech in relation to Korean stop contrasts according to age. In order to determine the relationship between F0 developmental patterns and age in child stop production, this study uses a quantitative acoustic model to examine the word-initial stop productions of 58 Korean monolingual children aged 20 to 47 months. The production experiment confirmed that VOT is useful for distinguishing fortis stops, but F0 is required for distinguishing between lenis and aspirated stops, and this tendency is significantly related to age. As F0 becomes a determinant acoustic parameter for articulatory distinction, the role of F0 in perceptual distinction was investigated through a perceptual identification test with the F0 continuum. Children were provided with selected lenis-aspirated pairs of images in which they would point to one or the other image in response to given synthetic sounds with different F0 values. This allowed us to observe how phonetic boundaries in the F0 dimension for aspirated stops change with age. A comparative analysis between children's production and perception of F0 indicates that articulatory skills depend on perceived F0 differences depending on the phonemic categories. Additionally, the analysis indicates that once F0 is acquired, VOT differentiation diminishes for distinction between lenis and aspirated stops, and this trade-off between VOT and F0 would occur around the age of 3 years. These findings suggest that phonemic categorization of lenis and aspirated stops should be processed in the F0 dimension and that phonemic processing in perceptual acoustic space is directly linked to phonetic discrimination between the non-fortis stops in production. This study provides experimental evidence for understanding a developmental trajectory of F0 as an acoustic cue for native phonological contrasts.

Table of Contents

3. PRODUCTION EXPERIMENT 41

4. PERCEPTION EXPERIMENTS 123

5. INTERRELATION BETWEEN CHILD SPEECH PERCEPTION AND PRODUCTION 175

LIST OF TABLES

LIST OF FIGURES

Chapter 1
Introduction

1.1. Significance of the study

Language learning begins with understanding speech sounds. Infants initially acquire human speech signals holistically in meaningful chunks, but soon realize that the chunks are sequences of sounds. During this process, they learn to recognize the phonetic parameters that make phonological contrasts in their native language and appropriately align speech signals to native phonetic boundaries. With these acquired phonetic parameters, they accept native contrasts but decline non-native contrasts, which ultimately allows them to interpret speech sounds (Werker & Tees, 1984). Thus, the acquisition of native phonetic parameters and the discrimination of native contrasts are closely related. This automatic process occurs early in language development. Through assessing children's early meaningful speech, we can examine what phonetic parameters they have acquired and what native contrasts they can distinguish. The present study deals with the acquisition of a native phonetic parameter and its role in the development of distinction of native contrasts.

Stops are the Manner of Articulation that children acquire first cross-linguistically, and Voice Onset Time (VOT), a universal acoustic parameter, serves to distinguish stops across languages (Lisker & Abramson, 1964). VOT represents the temporal relationship between the release of oral constriction and the vocal folds vibration. There are effectively three stop phonation types that differ in VOT, particularly when a stop sound occurs utterance-initially: lead-voice (voicing), short lag (voiceless unaspirated), and long lag (voiceless aspirated). Thus VOT is a key phonetic parameter through which to investigate children's developmental patterns regarding stop distinctions and has been used to study the process of stop acquisition in relation to the articulatory achievements of young children (e.g., Allen, 1985; Clumeck, Barton, Macken, & Huntington, 1981; Gandour *et al*., 1986; Kewly-Port & Preston, 1974; Macken & Barton, 1980a; Pan, 1994).

In spite of a number of studies on stop distinction in children, little research has been done on the acquisition of Korean stops. Korean has an unusual three-way contrast known as lenis, fortis, and aspirated, which are all pulmonic egressive voiceless stops. For this three-way distinction in Korean, VOT historically has been considered a useful tool, but the secondary acoustic measure, fundamental frequency (F0), is changing its role for Korean stop distinction. F0 is generally used as a prosodic term representing pitch contour in a sentence. However, in Korean stop distinction, F0 at vowel onset can be critical to identifying a preceding stop, since F0 values differ depending on what stop category precedes a vowel. In recent Seoul Korean, there have been reports of a tonogenetic sound change in progress from adult speakers' productions, in which F0 is becoming a primary acoustic cue for distinguishing non-fortis stop contrasts with the relative loss of the VOT difference among phrase-initial lenis and aspirated stops (Kang, 2014; Kang & Guion, 2008; Silva, 2006; Wright, 2007). Under this circumstance, the changing roles of VOT and F0 should be studied in relation to acquisition of Korean stop contrasts.

The present study broadly asks how young Korean children acquire the three-way stop distinction in relation to its acoustic parameter, F0. It is suggested that fortis stops are acquired first, followed by lenis and aspirated stops. In the acquisition of fortis stops, the role of VOT is determinant and is thus considered a universal mastery pattern (Kong, Beckman, & Edwards, 2011). However, the role of F0 has not been discussed in relation to the acquisition of lenis and aspirated stops, in spite of the fact that without an F0 difference, the series of three stops cannot be phonetically differentiated. As recent Seoul Korean is undergoing a tonogenetic sound change, F0 plays an important role in phonetic differentiation between the two voiceless stops. Therefore, it is worth investigating how F0 as a tonal contrast in the Korean stop system is acquired in young children’s linguistic performance.

In addition, while children’s articulation of Korean stop contrasts has been analyzed in the existing literature (e.g., M. -J. Kim & Pae, 2000; Y. -T. Kim 1992, 1996; Pae, 1994; Um, 1994), the perceptual capability of toddler speakers has not been previously studied in this regard. However, because language development occurs in the interaction between the perception and production systems, it is necessary to understand the mechanism of phonemic processing and acoustic implementation regarding stop contrasts.

Due to this dearth of analysis of children’s perceptual development, the questions of how phonemic processing occurs in children's acoustic space in terms of age and how F0-dimensional distinction affects the development of phonetic differentiation between lenis and aspirated stops remain unanswered. To answer these questions, the present project aims to investigate the perceptual capabilities of young Korean children and the interrelation between production and perceptual achievement with respect to stop distinction. Focusing on F0, the present study will show how the articulatory distinction between lenis and aspirated stops depends on children's perceptual abilities. Another goal of the study is to provide large-scale normative data on children's articulation and to present a pattern of native contrast development.

To pursue these goals, this study conducted experiments with 58 young children aged 20 to 47 months and provides a phonetic analysis of the data. The experiments consisted of a production test and a perception test. In the production test, to elicit their natural production of Korean stops, a picture-naming task was used with near-minimal triplets. In the perception test, a point-to-a-picture task was used with natural and synthesized stimuli. These experiments allow for a discussion of the phonetic accuracy of their stop production and how the phonemic categorizing process in perceptual acoustic space is related to articulatory distinction.

1.2. Outline of the dissertation

The dissertation is divided into six chapters. Chapter 2 reviews the background knowledge on the universal acoustic parameters for phonetic distinction across stop contrasts and synthesizes findings from the existing literature. In addition, it also assesses prior research on child language acquisition in terms of stop contrasts and on the two main acoustic parameters in Korean stop distinction, VOT and F0. Chapter 2 then presents the remaining issues to be answered and the specific research questions of this study. Chapter 3 deals with the production data of 24 adult and 58 children speakers of Standard Korean using a quantitative acoustic model. Chapter 4 presents a perception study, which consists of two different perception experiments with 48 children. The first experiment used natural stimuli and the second one used synthesized stimuli. The perceptual identification patterns by Korean young children are analyzed with a multilevel regression model. Chapter 5 compares the production data outcomes with the data on perceptual identification by children. Based on the interactive dynamics of speech production and perception, this chapter analyzes Korean young children's F0 acquisition pattern, which is represented by lenis-aspirated stop distinction. Chapter 6 summarizes the main findings of the production and perception experiments in the context of extant theoretical frameworks and concludes with suggestions for future research.

Chapter 2

Background and research questions

2.1. Stop contrasts

All languages have stop sounds, which are articulated by stopping one's airflow. Depending on what parameters are involved in the articulation of stop consonants, their phonetic sub-categorization differs, but most sounds are in fact produced using the same airstream mechanism—pulmonic egressive—in which the airflow from the lungs is blocked in the vocal tract. Stop sounds are acoustically distinguished by the active aspects of the sounds, which are realized in three temporal stages: closing, closure, and release. During closure, because it is difficult to maintain a sufficient amount of airflow for vocal fold vibration or for aspiration, voicing typically persists through part of the closure or disappears before the closure. This is why voiced stops are rarely found in most known languages, especially in the phrase-initial position. During release, the increased air pressure in the occluded vocal tract is released, resulting in stop release bursts; thus stop release bursts exhibit intrinsic variations, which depend on consonant Place of Articulation.

In the stop typology of the world’s languages, two-way stop contrasts are most common while four-way contrasts are rare. When stops have a two-way contrast, they are usually categorized in terms of voicing. Thus English has voiceless and voiced stops (/ba/ vs. /pa/) (Maddieson, 1984). If there is a three-way contrast, it is a general pattern that it is distinguished by voicing and the degree of aspiration. With voicing and aspiration, they are categorized as ‘voiced’, ‘voiceless unaspirated’, and ‘voiceless aspirated’ (Klatt, 1975; Ladefoged & Maddieson, 1996; Lisker & Abramson, 1964).

These acoustic features of stop contrasts can be phonetically measured by using Voice Onset Time (VOT) or fundamental frequency (F0). Having different ranges of these acoustic estimates, stop consonants can be phonetically differentiated and categorized as different phonemes in a stop system.

2.1.1. VOT

Voice Onset Time (VOT) refers to the duration of the period of time between the release of a plosive and the beginning of a vocal fold vibration and is an important parameter in identifying a stop category. VOT values generally fall into three different ranges, which correspond to voiced (leading), voiceless unaspirated (short), and voiceless aspirated (long) stops. Moreover, VOT values exhibit inherent variations depending on Place of Articulation and the following vowel. Velar and uvular stops have longer VOT values while bilabial stops have shorter VOT values (Cho & Ladefoged, 1999), and a following high vowel lengthens VOT whereas a following low vowel shortens it (Smith, 1978). VOT values also differ by contextual condition. For example, lexical stress can affect VOT values. In stressed syllables, VOT values are longer in the production of voiceless stops and the voicing contrast is distinguished by greater VOT differences in the stressed syllable (Lisker & Abramson, 1967).

VOT has been the most useful method to phonetically define stop categories. The crucial role of VOT in distinguishing stop consonants in many languages has been extensively studied (e.g., Cho & Ladefoged, 1999; Gandour *et al*., 1986; Klatt, 1975; Lisker & Abramson, 1964; Sundara, 2005). VOT ranges, which constitute a phonemic contrast in a language, are determined by the voicing vibrations during the production of stop sounds. If no vocal fold vibration occurs during closure but occurs shortly after in which case voicing starts after a delay, then we usually call it voiceless stop (i.e., /ta/ in English). There is voicing during closure and release in the production of fully voiced stops such as in Sindhi /da/, which has a negative VOT. In the production of aspirated stops such as /p^{h}a/ in Korean, there is no voicing during the closure, and at the moment of the stop release the opening of the vocal folds are still large, so it takes more time to start voicing vibration before the following vowel is produced. This case usually results in larger VOT values compared with voiceless unaspirated stops.

If a language has a two-way stop contrast, it is generally distinguished for voicing: one category is voiceless and the other one is voiced or voiceless aspirated, since in this way, VOT difference between the contrast can be maximal. When a language has a three-way contrast, it is highly possible to include one as a voiced stop category with two other voiceless stops that can be distinguished by the degree of aspiration (Cho & Ladefoged, 1999; Maddieson, 1984).

2.1.2. F0

Fundamental frequency (F0) is the lowest frequency of the signal. F0 is usually called 'pitch', which easily means the 'height of sound' as an auditory property in a universal way. This is of particular importance in defining intonation in prosodic phonetics, where it reasonably displays the movement of pitch. F0 can also indicate physical attributes, and F0 values show how stiff or slack the vocal folds are when producing, since the lowest component of the signals from the vocal fold vibrations means F0. F0 is a relative acoustic measure since F0 values differ depending on physiological factors such as the size of vocal folds or subglottal pressures. Due to this property, F0 values usually vary by gender or age. In spite of the relative nature of this phonetic measure, in the study of stop contrasts, F0 usually plays a supplementary role in distinguishing stop categories such as in F0 perturbation.

The correlation between F0 and the feature [voice] is well known; [-voice] of a consonant tends to raise the F0 of the following vowel, while [+voice] of a preceding consonant tends to lower the F0 of the following vowel (Haudricout, 1954; Hombert, Ohala, & Ewan, 1979; House & Fairbanks, 1953; Lehiste & Peterson, 1961). More specifically, the voicing characteristics of a preceding consonant only influence the early part of the following vowel. This co-occurrence of a high F0 and a preceding unvoiced consonant has been observed in many languages and is therefore considered a universal phonetic phenomenon.

However, other research has also reported that F0 can be a useful tool to identify valid phonetic differences between stop contrasts when other acoustic cues to voicing are very weak, and F0 perturbation is phonologized (e.g., Jessen & Roux, 2002; Jun, 1996; Keyser & Stevens, 2006; Kingston & Diehl, 1994). Despite the lack of research consensus on the correlation between F0 and unvoiced consonants, it has been observed in many languages that [-voice] of a consonant tends to raise F0 values at the early part of the following vowel.

2.2. Korean stop contrasts

The Korean obstruent system is unique in that it has three distinctive stop categories, and all are pulmonic egressive voiceless. The three different phonation types in Korean are called lenis (which is also called plain or lax), fortis (which is tense or reinforced), and aspirated. Each stop category occurs at three different places of articulation: labial, alveolar, and velar, as illustrated in the table below.

Table 2.1. Korean lexical consonant inventory.

	Labial	Coronal		Dorsal	Laryngeal
Lenis	**p (p)**	**t (t)**	tʃ (c)	**k (k)**	
Aspirated	**p^h(ph)**	**t^h(th)**	tʃ^h (ch)	**k^h (kh)**	h (h)
Fortis	**p’(pp)**	**t’(tt)**	tʃ’ (cc)	**k’ (kk)**	
Fricatives		s (s)			
		s’ (ss)			
Nasals	m (m)	n (n)		ŋ (ng)	
Liquid		r/l			

Note: The stop categories are bold. Yale Romanization in parentheses.

This kind of three-way distinction among stops is typologically rare. It is well-known that a two-way stop contrast is commonly distinguished for voicing, as in English which has voiced and voiceless homorganic stops (Maddieson, 1984). It is relatively common for a language to have two voiceless homorganic stops that are phonologically contrastive such as in Chinese or Hindi, but a language with more than three voiceless stop categories is typologically unique.

2.2.1. The phonology of Korean three-way stop distinction

Due to the unusual typology of Korean stops, researchers have attempted to distinguish the three-way contrast using feature-geometrical differences of the underlying forms of the three stops. For example, Kim (1965) proposed that Korean fortis stops should be underlyingly differentiated with the feature [+tense] or [+fortis] from lenis and aspirated stops with [-tense]. The feature [tense] means a tensing of the glottis and the supralaryngeal articulators such as the tongue or the lips. However, Kim’s argument has been revised by other researchers, as [tense] is not a universal feature in the world's languages. Using [tense] seems to have been improvised for a relatively small number of cases such as fortis stops in Korean, and the relation between underlying tones and [tense] cannot be well documented. Lombardi (1991) and Han (1992) illustrated the phonological differences in three stop series as follows without using [tense].

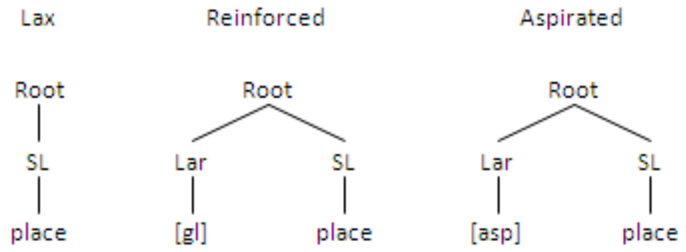

Figure 2.1. Phonological structures of Korean obstruents after Lombardi (1991).

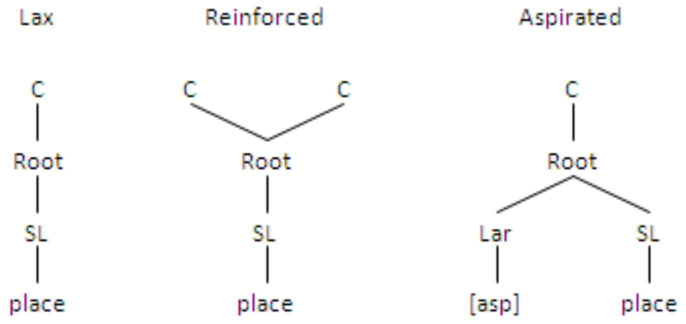

Figure 2.2. Phonological structures of Korean obstruents after Han (1992).

Lombardi and Han used different terms from Kim (i.e., lax for lenis and reinforced for fortis). The feature [gl] is equivalent to [+constricted glottis] and [asp] is equivalent to [+spread glottis] in the currently accepted feature geometry. Basically, their assumptions are similar in that aspirated stops have the laryngeal feature [aspirated], but Han added a time unit to Lombardi's structure. Han assumed that the time unit of the fortis stops should be double compared to those of the lenis or aspirated and that the laryngeal feature is not involved in this phonological structure. This analysis hypothesized that lenis and aspirated stops are singletons while fortis stops are underlyingly geminates (Ahn & Iverson, 2004; Avery & Idsardi, 2001; Han, 1996; Kim & Duanmu, 2004). Han's feature specification is motivated by important observations about the articulation of Korean stop contrasts. The duration of the stop closure has been considered an important finding in distinguishing Korean stop contrast. It is shortest for lenis and longest for fortis stops, while aspirated stops have intermediate stop closure. However, Cho and Keating (1999) showed that there is no significant temporal or spatial difference between the production of aspirated and fortis stops using an electropalatography (EPC).

About the laryngeal features adapted in this feature specification, the motivation to use [constricted] over [tense] is not explicitly explained. In addition, it is questionable whether Korean stops sound the same as other glottalic stops in other languages, which are phonologically represented as [+constricted glottis].

More recently, Kim and Duanmu (2004) proposed that lenis stops are underlyingly regular voiced stops but that they undergo devoicing word-initially, with the result that they are realized as voiceless in utterance-initial position. Their approach uses the feature [-stiff vocal folds] to account for the co-occurrence of lenis stops and low tone. The advantage of their framework would be that the new feature [tense], which is different from [tense] for vowels, does not necessarily play a role in defining fortis stops and that F0 perturbation can explain the co-occurrence of high tone-aspirated stops and low tone-lenis stops. Their feature specification captures the important laryngeal activity in the production of Korean stop contrasts. The fortis stops are produced with the highest glottal tension compared to the production of lenis or aspirated stops and this is because the glottis is raised for the production of fortis stops (Cho, Jun, & Ladefoged, 2002; Kim, Honda, & Maeda, 2005; Kim, Maeda, & Honda, 2010). They argue that in this approach, Korean phonology is not typologically unusual and is in accordance with universal principles of phonology. Kim and Duanmu's framework is appealing but has not yet been fully accepted.

More recently, Kochetov and Kang (2016) investigated palatolinguo contact in the production of Korean three-way stops and suggested that supralaryngeal characteristics enhance phonological analysis for Korean stop contrasts. Their feature specification is as follows.

Table 2.2. Feature specification for Korean stop contrasts from Kochetov and Kang (2016).

Korean (V_V) Features	Lenis	([slack])
	Fortis	[contricted], [stiff]
	Aspirated	[spread]

They argued that the feature [tense] does not predict temporal or spatial differences between fortis and aspirated stops but they are both underspecified for [tense], so this phonological analysis is not appropriate. However, the actual phonetic differences between fortis and aspirated stops can be presumably predicted by the use of [stiff] for fortis stops and [spread] for aspirated stops. The [stiff] means the tensing of the vocal folds and [spread] implies high subglottal pressure, and these phonological representations correspond with the acoustic implementation for fortis and aspirated stops (Kagaya, 1974). In addition, their palatolinguo results are equivalent to the results of the previous EPG studies (Cho & Keating, 2001; Shin, 1997), which indicates that their feature specification seems to reasonably reflect the articulatory characteristics of Korean stop contrasts.

2.2.2. Acoustics of Korean stop contrasts

A number of phonetic studies have extensively analyzed the acoustic properties of the Korean stop system. Most of them have proposed that Korean stops should be defined in terms of two phonetic parameters: Voice Onset Time (VOT) and fundamental frequency (F0). VOT plays a crucial role in distinguishing Korean stops as it does in the world's languages, in spite of the fact that voicing is not involved in stop contrasts in Korean. In particular, in word or phrase-initial position, Korean stops can be differentiated by VOT even though there are overlaps of VOT across the three categories (Cho, 1996; Han, 1996; Hardcastle, 1973; Han & Weitzman, 1970; Hirose, Lee, & Ushijima, 1974; C.-W. Kim 1965, 1970; M.-R. Kim, 1994; Y. Kim, 1995). Since Korean is a language with no lexical stress, VOT variation is not affected by stress patterns, while VOT does vary by lexical stress in other languages with lexical stress such as English (Klatt, 1975; Lisker & Abramson, 1967).

Korean stops that occur phrase-initially have three different ranges of VOT with some overlaps: aspirated stops have longer VOT values, fortis stops have shorter values, and lenis stops have intermediate values. Recent studies have consistently reported that VOT differences among the three-way stop contrast have been declining, particularly between aspirated and lenis stops, and have been replaced with other phonetic differences (Kang, 2014; Kang & Guion, 2008; Silva, 2006; Wright, 2007). Still, it is notable that fortis stops are distinctively separated from aspirated or lenis stops with their short VOT values, and the phonetic trade-off with F0 enhancement does not function for fortis stops (Kang, 2014).

Another key parameter in distinguishing Korean stops is F0. Many researchers (Cho, 1996; Han, 1996; Hardcastle, 1973; Han & Weitzman, 1970; Kagaya, 1974; Kim, 1994) have proposed that F0 is crucially correlated to the three-way stop distinction in Korean; in particular, F0 mainly contributes to identifying lenis stops and does not necessarily serve to distinguish between fortis and aspirated stops. The correlation between Korean obstruents and F0 has been addressed by Cho *et al*. (2002) and Kenstowicz and Park (2006). According to their experiments, vowels following fortis or aspirated consonants display higher F0 values than those that follow lenis consonants. This pitch pattern was understood from the perspective of F0 perturbation, in which such F0 differences are due to aerodynamic mechanisms (Haudricout, 1954; Hombert *et al*., 1979; House & Fairbanks, 1953). This phenomenon is regarded as a language-universal pattern motivated by physiological reasons. The F0 perturbation theory argues that the F0 difference in the target stop sound is affected by the [voice] feature; [−voice] creates higher vocal fold tension causing a higher F0, with the result that voiceless stops have higher F0 values and voiced stops have lower F0 values at vowel onset. However, Kingston and Diehl (1994) rejected the perspective that laryngeal activity is associated with the [voice] feature. They suggested that F0 variation can be found regardless of the absence or presence of [voice] and that F0 onset can be controlled to contrast stop sounds, which is supported by Jun's argument (1996). Jun proposed that F0 perturbation is a low level phonetic phenomenon. She argued that the effect of F0 perturbation on Korean intonation is limited only to an AP (Accentual Phrase) initial position, and F0 after lenis stops persists throughout the vowel in certain prosodic contexts. For this distinguishing characteristic in Korean, she suggested that F0 perturbation is phonologized at a prosodic level, which makes Korean stops phonologically contrastive. According to her argument, lenis stops are perceptually salient due to their lower F0 values. In addition, high tone (H) occurs with fortis and aspirated stops, while lenis stops do not appear with H. Thus, it is accepted that F0 perturbation as an automatic physical attribute cannot solely account for the correlation between F0 and the Korean stop distinction.

In addition to those two phonetic parameters, other aerodynamic factors pertaining to the three-way contrast have been studied by Han (1998) and Cho *et al.* (2002). Han investigated how the glottal configurations can be different in the three series of stops. He reported that the onset of the vowel after lenis stops has a breathy voice as indicated by positive H1-H2 (amplitude difference between the first harmonic and the second harmonic) values. H1-H2 indicates the tenseness of the vowel following the target stop consonant. Breathiness increases when the difference of H1-H2 is large, and glottal tension increases when the H2 value is larger than H1. Cho *et al*. examined the acoustic and aerodynamic characteristics of Korean consonants produced in Seoul and Cheju. In their research, they analyzed VOT, burst energy, F0, H1-H2, intraoral air pressure (P0), and airflow (U0) of the stops from their acquired recordings. They measured the maximum flow after release of the closure and peak oral pressure during closure. Their analysis indicates that, VOT, relative burst energy, F0, P0, and U0 are correlated with phonation type in word initial Korean stops, indicating that the three-way stop contrast can be differentiated by aerodynamic/physiological mechanisms, and their phonetic differences can be reinforced by multiple cues other than VOT or F0.

The assumption that among the two key parameters, F0 rather than VOT serves as a crucial cue for the Korean stop distinction has been proposed by Silva (2006), Wright (2007), and Kang (2014). Silva’s diachronic research on the phonetic differences between Korean stops showed that VOT differences between the lenis and aspirated stops have significantly decreased in young adult Korean speakers, resulting in F0 differences playing a determining role in distinguishing the three different manners. Along the same lines as Silva’s study, Wright (2007) also attempted to prove that the speech of the younger generation in Korea depends on F0 differences rather than on VOT differences. He conducted a phonetic analysis of production excised from natural speech by native

Korean speakers and showed that pitch difference is becoming important for younger people, which supports Silva's argument. In his perception test, he used resynthesized monosyllabic minimal pairs of varying F0 and VOT values. This experiment showed that pitch differences dominate VOT differences overall, but a correlation between age and the phonetic parameters was not found. Wright's work is distinguished in that he conducted a perception test with native Korean speakers; little research has focused on perception, despite the fact that extensive phonetic analyses of the Korean stop system have been conducted. More recently, Kang (2014) suggested that Korean stop contrasts are undergoing tonogenesis, which is the process of sound change, resulting in the loss of VOT differences and F0 enhancement among stop contrasts. She phonetically analyzed the corpus data from different generations and compared phonetic differences among the three series of stops. The analysis revealed a clear trend of tonogenetic sound change in progress in recent standard Seoul Korean, since F0 differences between aspirated and lenis stops were amplified and VOT differences were reduced in younger speakers.

Another notable perception study was done by Kim, Beddor, and Harrocks (2002). Their hypothesis was that in perceiving Korean stops, the contributions of consonantal and vocalic information are weighted differently. They excised the VOT portions of natural speech and used the vowel portions to determine whether the vowel alone could be a cue for the Korean stop contrast. In addition, they attached the extracted VOT portion of one phonation type to the vowel portion of another phonation type to include cross-spliced stimuli. The results suggest that the vocalic information with lenis onsets is sufficient to cue preceding lenis stops, while both the consonantal and vocalic information was required to identify fortis and aspirated stops. Thus, their findings indicate that F0 plays a determining role in distinguishing lenis stops in perception.

To summarize, the acoustic characteristics of Korean stop distinction have been studied extensively. In general, it has been suggested that both VOT and F0 serve as important phonetic parameters for the Korean three-way contrast even though their roles are different. However, there has been little research on the developmental process of the acquisition of such phonetic parameters. The next section will introduce the major views about the acquisition of phonemic categories and reported phonetic analysis of the Korean stop distinction in children.

2.3. Child acquisition of stop categories

As every spoken language has stop sounds in its consonant inventory, the acquisition of stop obstruents has been intensively studied in relation to their acoustic characteristics. Based on findings from adult speech, most acoustic characteristics involved in stop categories have been revealed. However, young children's learning abilities and their linguistic performance regarding stop contrasts have not been identified, and so understanding how they deal with language input and acquire certain native contrasts has been of interest to many researchers. Previous research on this topic has focused on whether the acquisition patterns of stop categories can be accounted for within a language-universal theoretical framework in order to determine whether there is empirical evidence for a theory of universal language development. Many studies have attempted to determine how young children acquire their native phonological contrasts and how they develop the relevant acoustic skills.

2.3.1. Phonological ordering

Regarding child language, Jacobson (1968) studied patterns in child phonology and asserted that there are universal patterns of language development in children. His proposal was mainly about the phonological ordering of acquiring phonemes, which he believed occurred in a strict order. The universal orders he predicted, regarding stop acquisition are as follows.

a. Vowels are acquired before consonants.
b. Stops are acquired before fricatives, and stops replace fricatives
  in the early stages.
c. Dental stops are acquired before velar stops.
d. Voiceless unaspirated stops are acquired before voiced stops.

His claim is based on the idea that there are marked and unmarked contrasts cross-linguistically. He believes that the innate knowledge of language guides the same universal acquisition of phonology, which is equivalent to the essence of Universal Grammar (UG). His markedness theory has been partly supported by many studies on child language acquisition (e.g., Amayreh & Dyson, 1998; Hua & Dodd, 2000; Stoel-Gammon, 1985; Templin, 1957), but other researchers have found counterexamples from various languages (e.g., Ferguson & Farwell, 1975; Macken & Barton, 1980a; Johnson & Reimers, 2010; Rice, 2005; Smith, 1973; Vihman, 1993, 1996). Jacobson's strict ordering is not observable in every language as there is no single order to acquisition, and different stop contrasts exist depending on the language—thus the variability of acquisition patterns cannot be accounted for within a single framework. Another problem of his theory is that his ordering does not take into account the variability of the surface forms of the same phoneme. Depending on the context, phonemes are realized differently. Thus position affects the acquisition of phonemes, but the relation of possible allophones to a phoneme is not included in the ordering (Johnson & Reimers, 2010). Thus, Jacobson's markedness theory provides a rough outline of child language acquisition but is not a theory of acquisition due to its problematic implications. Thus, the Jacobsonian view (as well as UG) has fallen out of favor, and recent research based on laboratory studies has emphasized the importance of speech input in the ambient language.

2.3.2. Theoretical grounds of language acquisition

The progress of linguistic development in early childhood depends on speech perception. Werker and her colleagues investigated how infants and young children deal with linguistic input and linked it to the adequate linguistic development growing up (e.g., Pegg & Werker, 1994; Polka & Werker, 1994; Werker & Pegg, 1992; Werker & Tees, 1984). Their extensive research was motivated from the observation that infants can easily discriminate non-native contrasts while adults have some difficulty doing so.

Throughout valid experiments with infants and toddlers, Werker and Pegg (1992) proposed the perceptual model described in Figure 2.3. They focused on the change in perceptual categorization in early childhood when a sufficient amount of native language input decreases infants' perceptual sensitivity so that they begin to lose the ability to easily discriminate non-native contrasts and their perceptual phonetic boundaries only distinguish native contrasts; their perceptual categories are reorganized to reflect linguistic experiences in their native language. At some later point in development, young children can use language-specific phonetic details to distinguish meanings, and this phonemic perception emerges around 19 months or later.

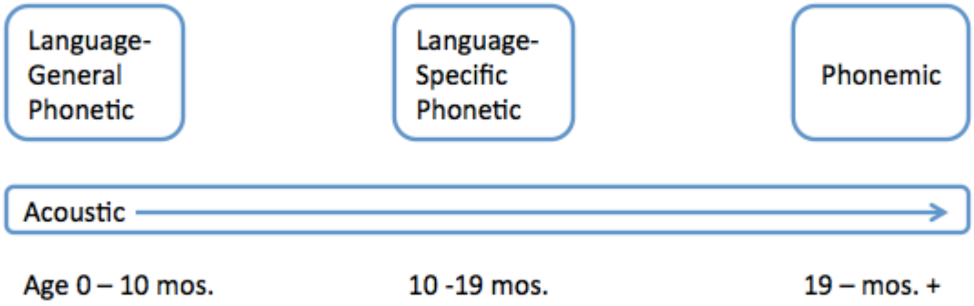

Figure 2.3. A four-factor model of speech perception. Adapted from Werker and Pegg 1992. (from Werker, 1995, p. 159)

The importance of linguistic input has been recalled in the Native Language Magnet (NLM) (Kuhl, 1994; 1998; 2000). The NLM accounts for how infants' perception develops language-specifically. According to the NLM, the prototype of a phonetic category in one native phonological system functions like a perceptual magnet, so that neighboring phones are effectively pulled into that category and assimilated to the prototype. The robust and good input will be the prototype and it operates as a magnet. The NLM proposes that the perceptual magnet is created in an acoustic space by speech input in a language-universal way at birth and begins to operate for vowels at the very least at around six months. Infants' perceptual distributional learning makes phonetic representations based on their language input and their phonetic boundaries undergo attunement in a language-appropriate way. This process occurs in infants' mental maps, and neural commitment is involved in this phonetic attunement, with the result that non-native contrasts are not easily acquired after acquiring a first language. Supporting the essence of the NLM, some studies have reported on the process of formation of native phonetic categories, demonstrating that younger children can discriminate non-native contrasts while older children cannot (e.g., Kuhl *et al.*, 1992; Polka & Werker, 1994; Werker & Tees, 1984).

Another major view on the receptive ability in language learning is the Perceptual Assimilation Model (PAM) (Best, 1993; 1995; Best, McRoberts, & Goodell, 2001; Best, McRoberts, & Sithole, 1988). PAM asserts that non-native sounds are assimilated into native sounds in various ways according to the perceived similarities and discrepancies between the non-native and native sounds. This model was developed from the basic idea of *the direct realist view* that perceivers directly pick up perceptual information through an integrated perceiving system and that perceptual attunement economizes perceptual input. The direct realist view proposes that listeners do not have to have innate knowledge of the vocal tract or acquired mental associations to perceive speech.

According to PAM, non-native contrasts will be assimilated into phonetically similar native contrasts, but the degree of phonetic similarity does matter in this assimilation process. However, PAM does not provide an explanation of how a number of non-native sound inputs are categorized into fine-grained native phonemic spaces. The basic assumption of the two major views is that perception is a fundamental and automatic part of the process of forming phonetic categories.

Consistent with the previous suggestions, the Theory of Adaptive Dispersion (TAD) proposes the characteristics of the native phonological systems with regard to the mechanism of forming vowel categories (Lindblom, 1986, 1998; Lindblom & Engstrand, 1989; Liljencrants & Lindblom, 1972). TAD provides theoretical grounds for an explanation of phonological changes in a native vowel system.

Lindblom and his co-workers highlighted two different characteristics of production and perception: maximal distinctiveness in perceptual information and economy of effort in vowel production, which are both needed in the phonemic configurations of vowel systems. They proposed that the compromising point of these two conflicting needs establishes one's native phonological system. Lindblom and Engstrand (1989) and Maddieson (1984) pointed out that maximal contrasts should be changed into sufficient contrasts and that vowel systems would evolve to have such sufficient perceptual contrasts. This means that vowel inventories have been attuned to consider these two factors. The fact that only small sets of vowel inventories are used in natural languages can be explained using TAD.

Psycholinguistic studies have attempted to provide evidence for TAD. For example, Johnson, Flemming, and Wright (1993) conducted a perception experiment with synthesized vowel sounds for English-speaking adult speakers, showing that the listeners preferred more dispersed vowel categories acoustically, which is called the “hyperspace effect.” They suggested that the logic of TAD means an active perceptual biasing pressure on the linguistic system (Johnson, 2000). Padgett (2001) expanded the basis of TAD to consonants, which is called “contrast dispersion,” suggesting that some allophonic shift in producing consonants would be motivated by maintaining a perceptually good contrast. Hence, it is assumed that having a perceptually sufficient distance in an acoustic space is essential for phonemic processing that takes place throughout children’s language development.

2.3.3. Stop acquisition in child speech and perception

It is assumed that production is guided by perceptual representations, since perceptual development precedes production in infants. Many laboratory studies of infants’ perception have used their discrimination ability as a criterion to judge their perceptual development. One of the studies reported that 1- and 4-month-olds could discriminate between [ba] and [pa] in English using a phonetic parameter, VOT, even when VOT values are changed between the categories for [b] and [p] (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Several studies by Swingley and his colleagues also demonstrated that young children are capable of discriminating the English stop contrast even though the primary goal of their research did not involve the stop contrast itself. For example, two studies by Swingley and Aslin (2000, 2002) tested children aged 14 to 23 months with stimuli that have a word-initial English stop sound and showed that their word recognition was influenced by the ‘correct’ onset sound. If the stop sound is replaced with another stop (e.g., *dog* to *tog*), it interferes with word recognition by young children. Using an eye-tracking method, Swingley, Pinto, and Fernald (1999) also exhibited that 24-month-olds are able to distinguish English stops (the pair of *ball* and *doll*). Their findings indicate that until the age of 24 months, children’s perceptual representations are established enough to discriminate their native phonetic contrasts, and the phonetic information is also stored with lexical meanings of words in the word learning process. The one thing to note is that all phonetic gestures are not acquired simultaneously in early language development. Rather, relatively salient and familiar phonetic features are first involved in forming phonetic categories so that perceptual representations continuously undergo phonological attunement throughout development (Logan, 1992; Metsala, 1997; Swingley, 2009; Walley, 1993).

Previously reported linguistic research on the acquisition of stop contrasts is biased on the production side. It has concentrated on the intrinsic aerodynamic mechanism for making stop distinctions. VOT is a successful measure representing the aerodynamic mechanism that pertains to phonetic stop differentiation, and it is a language-universal tool. Therefore, the VOT pattern in children’s production of stops has been observed by many researchers. For example, Kewley-Port and Preston (1974) showed that stops with short-lag VOT are acquired earlier than stops with long or voiced-lead VOTs. This is because articulating short-lag VOT stops is less demanding than the other two types of VOT stops, which require precise temporal control between the glottis opening and the oral constriction release. To produce voiced or long lag VOT stops, the gestures for maximum glottis opening and release of oral closure should be aligned and a significant difference of air pressure between the supra-glottis and the sub-glottis should be made so that the vocal folds start to vibrate during the closure (Keating, Linker, & Huffman, 1983; Westbury, 1983). However, short lag VOT stops are produced by the glottis opening during oral closure, which does not require precise temporal coordination (Kong *et al.* 2011). Due to aerodynamic and physiological reasons, it is predicted that short lag stops are acquired earlier cross-linguistically. With this prediction in mind, phonetic analyses of children's production in various languages are summarized in the table below.

Table 2.3. Summary of earlier studies from Kong (2009, p.8).

Language (source)	Age and VOT pattern
	asp. vs. unasp.
English (Kewly-Port & Preston, 1974)	Separate VOT peak for /t^h/ at 75 week
English (Macken & 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 & 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

Since these languages have different stop contrasts, the correlation between age and VOT variation differs depending on the language. In English and Cantonese, both languages that have aspirated and unaspirated stop contrasts, infants begin to differentiate the stop contrast by using longer VOT. With a three-way distinction in Thai, children start to show significantly different VOTs for aspirated vs. unaspirated around the age of 3. For languages with more than three stop series that are phonologically contrastive such as Hindi, 6-year old children do not yet differentiate VOT values. Thus, the development of stop contrasts in children has been studied in many languages, and it is suggested that a universal phonetic parameter can create language-specific phonetic rules as shown in Pierrehumbert (1980) and Keating (1985).

In the case of Korean, a few phonetic analyses of children's speech in terms of VOT and F0 have been reported. Han's (1998) study attempted to reveal the phonetic characteristics of the production of Korean stops by children at the age of 10. She recorded the production of isolated words and words in sentence contexts by the children. In isolated words the production of children showed adult-like patterns, while in sentence contexts the three stop manners were not consistently differentiated by VOT and F0. Her finding suggests that children at the age of 10 have not yet established adult-like patterns of articulation. In addition to Han's study, Kang (1998) tried to investigate the production of much younger children aged 2;8 (years;months). She measured the phonetic properties of Korean obstruents produced by the 2;8-aged children and reported that VOTs in aspirated and lenis stops were overlapping and that adult-like F0 differences were not found at that age. Kim (1999) also examined VOT ranges by four child speakers. His target age was 5 to 7 years, and his analysis showed that VOT values overlapped especially between lenis and aspirated stops, which was proved with 2;8-aged children in Kang's work. With this result, he concluded that VOT differences cannot distinguish between the three different stop categories and that the three phonation types need to be differentiated by other acoustic parameters such as stop closure duration and amplitude differences. While the implications of his study appear useful, his study tested a small number of subjects, which limits the generalizability of his findings.

Kim and Pae (2005) attempted to provide normative data for children's speech. They recruited a large body of children aged 2;6 to 6;0 to test them with 19 Korean consonants within various contexts. In the presence of the experimenter, a child participant was asked to name the given picture. The experimenter then assessed the child's articulation. Through this experiment, they revealed that the percentage of correct consonants for late 2 years of age was around 60%, but after the age of 4 the percentage was over 90%. Their results also showed that fortis stops are acquired earlier than lenis or aspirated stops, showing that initial /p'/ and /t'/ were correctly produced by 95% of children before 2;6. Their findings are contrary to the report by Kim (1996) that lenis consonants were acquired earlier than fortis consonants but are in accordance with Bortolini *et al.* (1995) in that long lag stops are more difficult to acquire since they require complex intrinsic laryngeal muscle activities to produce the time distance between articulation and real vocalization. Moreover, they reported that stop sounds are acquired earlier than fricatives, so in many cases in children's speech under age 6, fricatives are actually replaced with corresponding stop sounds. Their findings partly support Jacobson's acquisition order, however, given that the results are not from acoustic analysis but from observational judgment, it may be difficult to generalize the observed patterns.

Lee and Iverson (2008) conducted a quantitative analysis in an attempt to reveal the relation of VOT and F0 to three stop series in children's speech. They recorded 30 children's (aged 5;0 and 10;0) speech in an experimental setting. Their findings indicate that sex and gender differences are correlated with VOT and F0 values. The children tended to use temporal coordination that presented as VOT differences distinguishing three stop manners, and VOT values in lenis and aspirated stops did not overlap. This result is surprising because VOT overlaps in young adults' speech have been consistently reported (Kang & Guion, 2008; Silva, 2006). They concluded that adult-like Korean stop articulation could be acquired as early as 5 years of age even though their VOT values are greater at that age than those of adult speakers. Later, Lee and Iverson's work expanded to an acoustic examination of Korean stops produced by Korean-English bilingual children aged 5;0 and 10;0 (Lee & Iverson, 2011). In this study, they assumed that the two different languages mutually influence each other, and their results indicate that bilingual children aged 5 tend to show longer VOTs for Korean stop production from the influence of English. Their articulation of Korean stops generally depended on VOT, not F0. However, children at the age of 10 used both VOT and F0 to distinguish all stop series in their production. With these findings, they suggested that bilingual children's articulations are different from Korean monolinguals' and that depending on age, VOT and F0 patterns vary in the production of bilinguals.

The most recent and systematic report about children's stop production can be found in Kim's work (2008). She focused on the phonetic development of word-initial Korean obstruents in young Korean children. As described, research on the process of acquisition of the stop system by infants or young children is extremely limited. Considering this, Kim tried to illustrate the developmental pattern in the acquisition of Korean word-initial stops with a large body of data from 40 young children aged 2;6 to 4;0. For a comparative analysis between age groups, she had four age groups with a 6-month interval between groups, so that she had groups of 2;6, 3;0, 3;6, and 4;0. She used near-minimal triplets in an experimental setting and obtained VOTs, F0s, and H1-H2 from the recorded speech. She reported that except for the 3-year-olds, the other three groups showed significantly short VOT values for fortis stops with no significant difference between aspirated and lenis stops. The 3-year-olds showed significant VOT differences for all stop categories. Except for that group, VOT overlaps between fortis and lenis or aspirated stops at all places of articulation were found overall. With respect to F0, all age groups demonstrated significantly higher F0s for fortis and aspirated stops, but only the 3;6 age group successfully differentiated fortis from aspirated stops with F0 differences. The other three groups failed to contrast the other two stop categories: lenis vs. aspirated. With regard to H1-H2, the results indicate that H1-H2 tends to co-vary with VOT (e.g., Ahn, 1999); there was no significant differentiation between aspirated and lenis stops, and observable overlaps between the two stop categories were found for all age groups. Another contribution of her work is that she provided native adult speakers' judgments on the children's production. The collected data were classified by two native adult speakers who judged the phonemic correctness of each token. The results show that most of the lenis stops produced by all age groups were misclassified, which implies that lenis stops are acquired last. Hence, her research suggests that F0 differentiation emerges as early as 2;6 but is not fully established as adult-like articulation, which is in accordance with Jun's (1996) argument that F0 is a language-specific phonological component. This is a comprehensive and systematic examination of multiple phonetic cues for the Korean stop contrast. However, her results between groups are inconsistent, which makes it unadvisable to draw conclusions about apparent developmental patterns in children's stop distinction. This inconsistency could be because the grouping of subjects into 6-month intervals is not well aligned with the relevant developmental patterns. Furthermore, language development varies from one child to the next, which makes it difficult to group children by age and generalize a pattern from it; Kim may have overestimated the correlation between age and linguistic ability. Another weakness of this work is that it lacks an important part: perception. The developmental process is generally studied in the interrelation between the production and perception systems, but Kim's research provides an analysis of only children's production.

Focusing on the order of mastery among the three stop categories by children, Kong (2009) and Kong *et al.* (2011) revealed the relation of multiple phonetic properties of Korean stops to perception. Their hypothesis was based on the idea that there is a certain order to mastering the laryngeal features of Korean stop consonants, and this order affects perception by adults too. The observable differences from Kim's (2008) work would be that their target ages were widened from 2;0 to 5;11 and they used a repetition task. The child participants were asked to repeat an audio presentation with a picture. Through this repetition task, VOT, F0, and H1-H2 from the production of stops were measured. The analysis indicated that VOT cannot be used to differentiate aspirated from lenis stops, which has been reported in adult speech (Kang & Guion 2008; Silva, 2006). Instead, F0 served as a critical cue to distinguish lenis from aspirated stops. The findings of this phonetic analysis are supported by the results of a perception test. Adult listeners are dominantly affected by VOT in correctly identifying fortis stops. Considering the role of VOT in production and perception, the pattern of acquiring and mastering fortis stops earlier than the other two stops should be universal, not exceptional. This finding is in accordance with Kim and Pae's (2005) proposal that fortis stops emerge earlier than lenis stops. The weakness of Kong and Kong *et al.*'s work is that they used a repetition task to elicit children's production. Through repetition with an audio-visual presentation, children can learn or be trained to articulate a given sound, which raises the question of whether such trained production can really provide insight into their developmental stages. In addition, if their production was not correct/adult-like, it is unclear whether (1) they perceived the phonetic distinction among the stop categories but their articulatory immaturity was an obstacle to correct production, or (2) they could not recognize perceptual differences in the first place and they articulated what they perceived. Given these shortcomings, Kong and her colleagues' work should have explained how their collected data reliably represent the child mastery pattern.

2.4. Remaining issues

Previous research on stop distinction by young children mainly observed phonetic properties in production. Cross-linguistically, VOT plays a key role in distinguishing homorganic stops that are phonologically contrastive, so many studies have attempted to show children's VOT patterns in order to discuss their linguistic developmental stages. However, in the case of the Korean stop contrast, VOT cannot solely account for the developmental pattern of young children's production, so another crucial acoustic property, which is presented as F0, has been analyzed with VOT. A few large-scale studies have exhibited normative data of children's speech in early stages to investigate the developmental process regarding stop distinction, but there are remaining problems.

First, the biggest problem of previously reported studies is that they all lack the study of perception. While the acoustic cues for the Korean stop contrast have been analyzed on the production side, children's perceptual capability has not been discussed. In order to complete a story of language development, the relationship between the production and perception systems should be dealt with within the same framework. A tight link between perception and production is emphasized in Kuhl's NLM model. Perceptual sensitivity guides production in young children, so children's perceptual capacity can tell us about the development of production. Without a perceptual study, we are unable to determine what is responsible for the distinctive production of three stop manners. For example, if a certain age group shows relatively poor articulatory distinction, researchers need to understand whether the poor articulatory distinction is caused by clumsy articulatory control or by imperfect perceptual differentiation. The degree of accuracy in their articulation should depend on what phonetic/phonemic categories children have established and on what phonetic boundaries there are. This means that perceptual ability is responsible for achievement in production. Phonetic distinction develops in the interaction between the production and perception systems, so it is worth focusing on the perceptual side to establish a reliable developmental pattern for distinguishing stop categories.

Second, the role of F0 in establishing phonemic stop categories in young children's acoustic space has not been clearly illustrated. Most previous studies focused on phonetic analysis with VOT and F0, but made assumptions about the relation between VOT and the articulatory stop distinction. As F0 becomes more determinant even in adults' speech in the process of tonogenetic sound change, it is worth investigating the changing role of F0 in stop distinction in toddlers' speech. We can ask how F0 affects the formation of the phonemic categories regarding three stop manners and at what age F0 is dominantly involved in phonetic attunement. Thus, the effect of F0 on the acquisition of Korean stop contrasts should be demonstrated.

Third, normative data on children's natural speech are extremely limited, especially at very early stages in terms of language production. A few studies have provided large amounts of data from children's natural speech in an experimental setting (e.g., Kim, 2008; Lee & Iverson, 2008), but the youngest age group was 2;6 in Kim's work. The other studies that dealt with younger children did not conduct any phonetic/acoustic analysis. Instead, their methods were essentially naïve observation. Therefore, the early stage in language development where children start producing meaningful early speech (which is different from babbling) has not been reported on from a phonetic perspective. In particular, as previous studies noted (Kim, 1996; Kim & Pae, 2005; Pae, 1994), it is suggested that F0 distinction emerges around age 2 in Korean. Thus it would be worth investigating stop contrasts by children at age 2 in order to provide a more comprehensive explanation about early acquisition stages with reliable phonetic evidence.

Last but not least, tracking the developmental process of stop distinction in relation to age has not been attempted. That is, the relationship between toddlers' age and their phonetic stop distinction has not been clearly explained. This is because studies have concentrated on production acoustic analysis, so a developmental trajectory of stop distinction in the dynamics of production and perception remains vague. The second reason is that studies have not dealt with age as a continuous variable. Rather, the fundamental premises of their analysis were that age is positively correlated with language development and that children in the same age group share the same linguistic developmental stages, so they focused on showing analysis between different age groups even though each child had different linguistic competence. So their phonetic analysis is not enough to generally support or confirm Jacobson's universal theory. If age is considered a continuous variable in analysis, it is possible to examine how significantly correlations between age and each phonetic cue exist. With this treatment, the overall developmental pattern with regard to distinction for phonologically contrastive stop categories with VOT or F0 can be investigated.

2.5. Goals of research

As described in section 2.4, there is a large gap in previous research on the phonetic development of Korean stop distinction. In spite of the fact that perceptual learning is crucial in the development of phonetic and phonemic categories to make phonological contrasts as the NLM asserts, the interactive dynamics between production and perceptual systems in developing an acoustic parameter has not been fully investigated, largely due to practical difficulties in collecting data on young children in an experimental setting. Previous research has concentrated only on the acoustic analysis of the production of three different phonation types, which lacks the important link between perception and production. As a result, one side of the story remains in question. To fill this gap, the present research will examine the progression of perceptual development through the establishment of phonemic categories in acoustic space.

More specifically, as F0 values are needed to distinguish among the Korean three-way stop contrast, how F0 as an acoustic parameter comes to establish phonetic boundaries for the three series of stops in young children will be investigated. Considering the phonetic trade-off between VOT and F0 in distinguishing stop contrasts in the recent tonogenetic change of Seoul Korean, it is more important to track how F0 is acquired and functions in the distinction for the laryngeal contrasts in young children's linguistic performance. The interactive developmental patterns of F0 between production and perception will be investigated through two different experiments: a production test and a perception test. The research questions that will be answered are as follows.

a. What phonetic properties can be found in production of Korean three-way stop contrast by toddler speakers? How do the phonetic realizations of the stop contrasts differ according to children's age?
b. What is a perceptually salient phonation type? Does perceptual salience depend on phonetic details, and if so, which ones?
c. How does phonemic categorization in the F0 dimension develop with age? How does this process affect perceptual and articulatory discriminatory ability?

The answers for those questions can contribute to the investigation of the interactive development of toddlers’ linguistic competence that is represented as perceptual discrimination and acoustic implementation. A comparative study of both sides of linguistic development can also provide empirical evidence for the acquisition of native phonological contrasts in toddlers.

Chapter 3

Production experiment

The main purpose of this study is to uncover a clear trajectory of the development of F0, which is a crucial acoustic cue for Korean stop distinction in child production and perception. To investigate the production system of young child speakers, it is necessary to conduct phonetic analysis of children's actual articulations involving the related acoustic parameters, F0 and VOT. A production experiment was designed and conducted to provide a phonetic standard of the current stage of toddlers' stop category learning in a multidimensional acoustic space. The expected result of this experiment was a clear relationship between age and F0 development, which can be represented as significant F0 differentiation across different stop categories, as well as evidence that toddlers acquire VOT prior to F0, which can be suggested by more mature productions of fortis stops.

3.1. Participants

A total of 58 children varying in age from 1;8 to 3;11 (years;months) participated in the production experiment. Participants' gender was not considered in recruitment (Table 3.1). All participants were native Korean monolinguals without hearing or speaking disorders, as reported by their parents/guardians. The participants were recruited from two daycare centers in Seoul, Korea. The experiment took about one hour and was conducted in a quiet room in a daycare center library. Before the session started, each child was given instructions about what they are doing and they were supposed to do by their teacher and the experimenter. Each child was accompanied by a teacher to the experiment room, but all child participants performed very well without their teacher once the experiment started. In the experiment room, the experimenter's laptop, speakers, and a microphone (omnidirectional YETI) were set on the table; on the floor, toys and books were available in case a participant felt they were not ready, so they could play with them until they felt comfortable enough to join the experiment. Each child was assisted and encouraged to freely and comfortably join the experiment by their caregiver and the experimenter. They were also fully informed that they should express any uncomfortable feeling whenever they were not ready or did not want to participate, so the experiment could be stopped right away. As a result, all 58 child participants were very willing to participate in the experiment and gave good quality natural productions of three different Korean stop categories.

Table 3.1. Child participant information.

Age group	Male	Female	Total
1;8-2;0 (20-24 mos.)	2	2	4
2;1-2;6 (25-30 mos.)	5	6	11
2;7-3;0 (31-36 mos.)	6	8	14
3;1-3;6 (37-42 mos.)	7	8	15
3;7-3;11 (43-47 mos.)	3	11	14
Total	23	35	58

Table 3.1 shows the five different age groups of young child participants for the sake of convenience, but further explanation about the population is necessary. The participants' actual birth dates were all different, since there were no limitations on age during recruitment. Therefore, age is the most crucial independent predictor in the analysis for child phonetic development.

The number of participants in the youngest age group (1;8 - 2;0) is the smallest, as children under 2 years of age are sometimes simply too young to produce robust sound tokens (which consist of real words). Their productions of the tokens used for this experiment were more likely to be produced in a standard way, not with their own pronunciation. In spite of the unbalanced numbers of participants in each age group, including variation in age is clearly a beneficial way to determine any effect of age on phonetic development in regard to Korean stop distinction.

3.2. Methods

A picture-naming task was used for the production experiment. Two sets of nine words which include three different stop types (lenis /p, t, k/, fortis /p', t', k'/, and aspirated /p^h, t^h, k^h/), were carefully selected so that each word was able to be represented with a describable picture for which even the youngest child would know the name. All pictures were provided in randomized order, and each picture was shown three times with the restriction that no picture followed itself successively. The words used are as follows:

Table 3.2. Word list produced by child participants.

Target phoneme		Target word	English gloss	Target word	English gloss
Labial	p	/paŋ/	room	/pɑl/	foot
	p’	/p’aŋ/	bread	/p’alæ/	laundry
	p^h	/p^{h}al/	arm	/p^{h}at/	red beans
Alveolar	t	/tal/	the moon	/tak/	rooster
	t’	/t’al/	daughter	/t’ʌk/	rice cake
	t^h	/t^{h}ajo/	cartoon character	/t^{h}ʌk/	chin
Velar	k	/kɔŋ/	ball	/kawi/	scissors
	k’	/k’ɔt/	flowers	/k’atʃi/	magpie
	k^h	/k^{h}ɔŋ/	beans	/k^{h}amɛra/	camera

The sets of words used include lenis, fortis, and aspirated word-initial stops, and the stop category varies by Place of Articulation (henceforth, POA) including labial, alveolar, and velar stops. In the presence of the experimenter, the participant was asked to name each given picture. The participant’s responses were elicited to name each given picture with the question “*mwuetici?*” meaning ‘what is this?’ by the experimenter. If a participant had difficulty, the experimenter assisted them in naming the given picture. An omni-directional YETI microphone was used for the recording and was placed at a distance of 3-15cm from the participant’s mouth. The whole session was recorded and digitized on Praat version 5.3.01 (Boersma, 2001) set up on a personal laptop at a 22050 Hz sampling rate and an 11025 Hz filter rate with 344 Hz of bandwidth. The recordings were conducted in a quiet room in a daycare center library and saved as WAV files.

3.3. Procedure

Target words were extracted from the recordings and the word-initial stops were analyzed using Praat version 5.4.18 (Boersma & Weenink, 2015). Through this experiment, the goal was to obtain a total of 3,132 tokens (18 words × 3 repetitions × 58 speakers). However, the number of tokens from each participant varies due to vocabulary and other practical issues such as quiet voices, amount of noise, too creaky or breathy voice, or other unanalyzable phonetic features. As a result, nearly 2,300 tokens were selected as appropriate for phonetic analysis, and their VOT and F0 values were obtained. The measurement of VOT was made from spectrograms and waveforms. Mainly based on spectrograms, F0 at the following vowel onset was estimated at the onset of the second formant. These measures were automatically calculated in Praat.

3.4. Control group

3.4.1. Adult participants and procedure

To provide a phonetic standard for Korean stop contrasts for comparison with children's productions, this study collected productions of the same target words from a control group of 24 native adult speakers of Korean. The control group for production consisted of 10 male and 14 female adult speakers, whose age range was 26-30 (born in 1980s), and who were all from Seoul and living in Seoul, Korea. No hearing or speaking difficulties were reported. All participants were paid $20 for a half-hour session.

The recordings were conducted in a very similar setting as in the child experiment. In a quiet room in a college library, each speaker was asked to articulate the given word list within a carrier sentence "*Tasi* [target word] *haseyo*," meaning 'say [target word] again.' A YETI omni-directional microphone was used for recording the sessions. The signals were recorded and digitized on Praat version 5.3.01 set up on a personal laptop at a 22050 Hz sampling rate and a 11025 Hz filter rate with 344 Hz of bandwidth. From this recording, a total of 540 tokens from male speakers (18 words × 3 repetitions × 10 speakers) and 756 tokens from female speakers (18 words × 3 repetitions × 14 speakers) were obtained. As in the experiment with children, the same phonetic estimation was conducted with the obtained sound tokens. VOT and F0 values of every token were extracted and used for phonetic analysis.

3.5. Analysis method

The raw data acquired from the child participants' actual productions of the target words were acoustically and statistically analyzed. The phonetic parameters used are VOT and F0. The statistical analysis used three different statistical models and showed the degree of significance of the hypotheses.

3.5.1. Acoustic measures

3.5.1.1. VOT

VOT is a key acoustic parameter that determines acoustic characteristics of Korean stop contrasts as it does in other world languages, even though there is overlap in VOT across the three stop categories in Korean (Cho, 1996; Han, 1996; Han & Weitzman, 1970; Hardcastle, 1973; Hirose *et al.*, 1974; C. -W. Kim 1965, 1970; M. -R. Kim, 1994; Y. Kim, 1995). As tonogenesis seems to be in progress in Korean, the importance of VOT has been somewhat underestimated, especially in distinguishing between lenis and aspirated stops (Kang, 2014; Silva, 2006; Wright, 2007). However, the role of VOT is still crucial in distinguishing fortis stops in Korean by the distinctively small degree of aspiration, which is represented by its short VOT value. VOT is defined as the duration of the period of time between the release of a plosive and the beginning of vocal fold vibration, so with a spectrogram and waveform, it is calculated from the very beginning of aspiration to the appearance of clear formant bars that indicate the sounds of a following vowel.

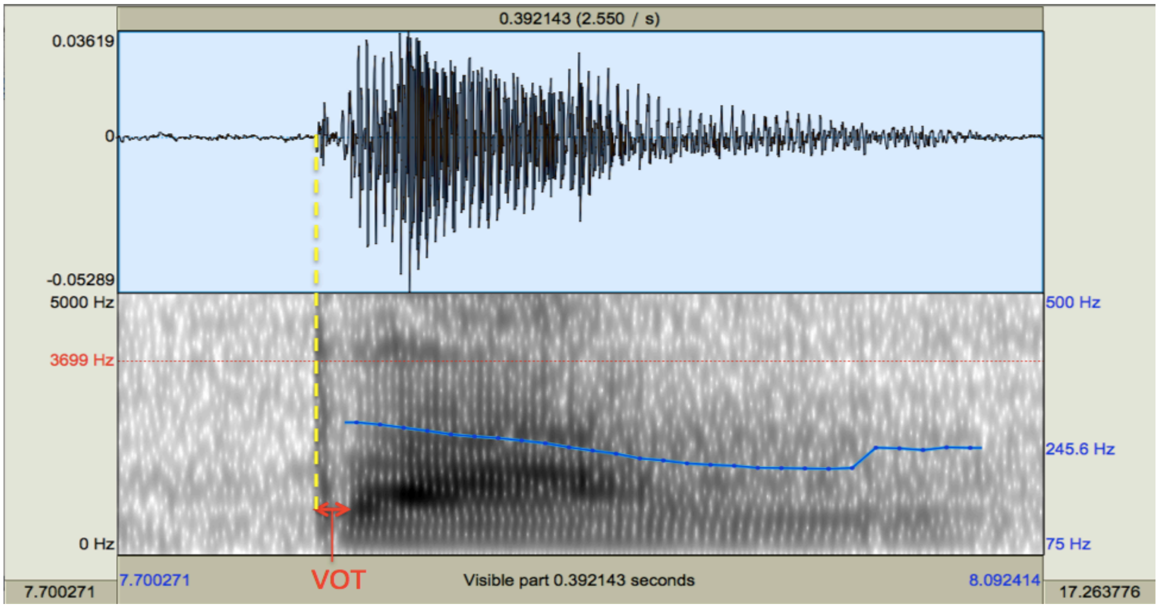

Figure 3.1. The calculation of VOT in the production of /p'aŋ/ ('bread') by *childid_14*, aged 40 months. The yellow dashed line indicates the beginning of aspiration for fortis /p'/. The red arrow indicates the calculated VOT (ms) in this case.

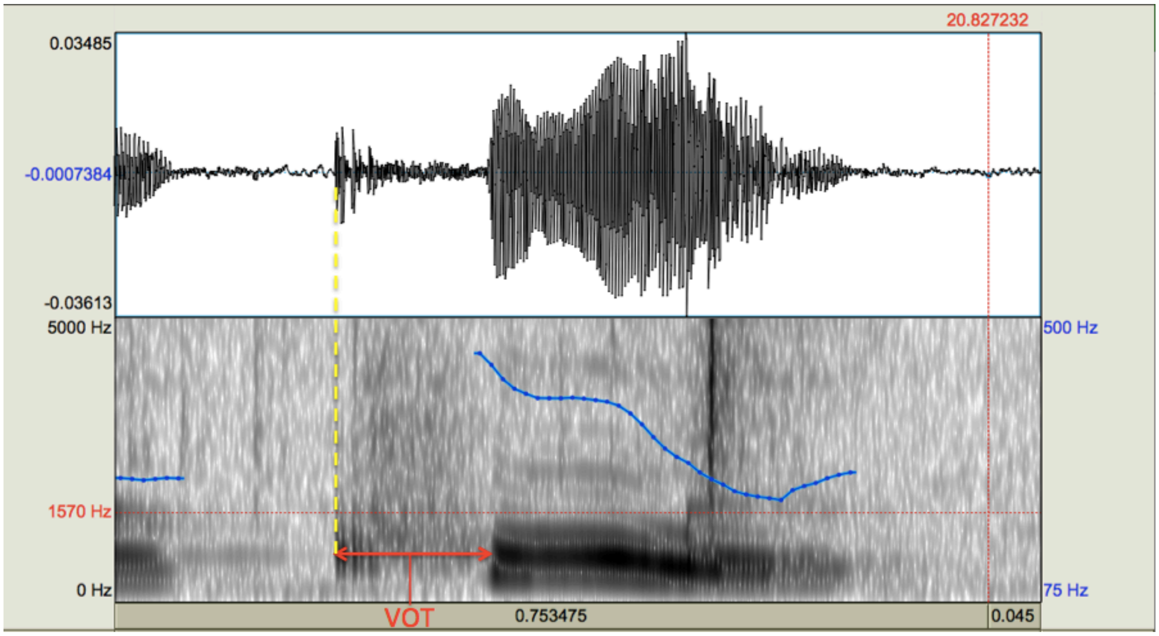

Figure 3.2. The calculation of VOT in the production of /k^hɔŋ/ ('beans') by *childid_14,* aged 40 months. The yellow dashed line indicates the beginning of aspiration for aspirated /k^h/. The red arrow indicates the calculated VOT (ms) in this case.

As the two figures above represent, VOT varies depending on the manner of the utterance-initially stop: fortis stops have a much shorter VOT, while aspirated stops have relatively longer VOT.

3.5.1.2. F0

The correlation between pitch pattern and obstruent has been understood through the perspective of F0 perturbation, in which the F0 value in the target stop sound is affected by the [voice] feature. In Korean, this is realized as a low-level phonetic phenomenon, so the effect of F0 perturbation is limited to AP initial position (Jun, 1996). As Korean undergoes tonogenesis and VOT differentiation between lenis and aspirated stops becomes less salient, F0 has been regarded as the most crucial acoustic parameter to differentiate these two stop categories (Kang, 2014; Silva, 2006; Wright, 2007). Considering the determinant role of F0 in distinguishing between Korean stop contrasts with the recent trend of tonogenesis, it is necessary to compare the F0 values of the three different stop categories, to reveal any phonetic trade-off between VOT and F0 values and its consequences on the acquisition pattern of young children.

To measure F0 as an acoustic cue for identifying a certain stop manner, it is important to get the values at the vowel onset right after the target stop sound, as indicated by the onset of the second formant. Using the spectrogram, F0 at the vowel onset was automatically measured in Praat.

3.5.2. Statistical analysis

This section introduces what statistical methods were used for the production data and explains why those methods were applied. This study was designed in order to reveal the developmental patterns of the crucial acoustic parameter, F0, according to children's age. Since the Korean stop system needs to be defined in a multi-parametric acoustic space, it is important to use those key parameters as dependent variables that are directly related to the articulatory variables in the analysis. As previously mentioned, the main assumption is that age is a key factor affecting the speech processing stages. Therefore, the acquisition and use of VOT and F0 should be primarily studied with age as a factor, so the degree of significance of the predictable effects of factors can be determined by statistical methods. The obtained phonetic estimates include a large amount of VOT and F0 values, and these numbers should be organized in optimally appropriate ways. For this process, some useful statistical methods are required.

In order to figure out what statistical design is appropriate for the data, it is necessary to consider and narrow down the primary concern of this study. First, through the analysis, it should be determined whether each child's productions of the three different phonation types are phonetically distinguished or not. To address this, all of their performances in pronouncing each stop category were analyzed with the two representative parameters, VOT and F0. Therefore, a two-sample t-test (Snedecor & Cochran, 1989) was conducted to compare VOT and F0 values in three pairs of stop groups (lenis-fortis/lenis-aspirated/fortis-aspirated).

The second concern is determining how appropriate and accurate the children's VOT and F0 differentiation are. Since the children's production data includes diverse production patterns with outliers, comparison with the adult production data can identify mature productions by the children. This study has a control group consisting of Korean monolingual adult speakers who provided normative production data. The best way to compare the child data to the control group's data would be another two-sample t-test. With these parametric statistical methods, I can make a judgment about the degree of similarity or difference between child and adult speech. This can help us figure out the efficacy of each child speaker's linguistic performances in distinguishing between Korean stops.

Using t-tests, it is possible to measure each child participant's linguistic ability, but it still hardly proves the effect of age on linguistic performance with a clear correlation. To draw a clear relationship between age and phonetic development in regard to VOT and F0 differentiation for stop contrasts, a simple linear regression model was used. In this analysis, the only independent variable is child age, and performance as represented by VOT and F0 values is the dependent variable.

Lastly, to find the statistical significance of interactions between predictor variables in child data, a mixed-effects linear regression model was used (Baayen, Davidson, & Bates, 2008). The hierarchical linear regression model is designed based on multi-level data and it focuses on the children's production patterns in relation to factors. This multi-level analysis can provide a more accurate prediction about F0/VOT development and articulatory distinction by children.

3.5.2.1. The two-sample t-test

A fundamental assumption to test is that child speakers recognize phonetic differences between stop categories and that the perceived phonetic differences can be represented by the articulatory distinction. A t-test is basically designed to compare values, and this characteristic makes it useful for the comparison of one speaker's production in two different contexts. Through a t-test analysis, it is possible to compare one speaker's fortis production with lenis or aspirated production since it can determine whether those comparisons are statistically similar or not. It should be noted that this does not tell us whether the observed difference or similarity is acoustically appropriate or not. However, knowing the degree of similarity or difference in productions of two different stop groups is valuable by itself.

A two-way t-test can compare VOT or F0 values across three different stop categories produced by one child speaker. Therefore, a two-way t-test was conducted for each child six times (a total of 348 = 58 participants × 3 different pairs × VOT or F0); I compared VOT and F0 values respectively in (1) lenis and fortis stops, (2) lenis and aspirated stops, and (3) fortis and aspirated stops. If this model judges the compared values statistically different enough ($p < 0.05$), the difference in VOT or F0 can be considered a significant phonetic distinction between two stop categories.

Through this statistical method, it is possible to judge the phonetic similarity or difference of VOT or F0 in the productions of children. The next question is whether the phonetic difference in VOT or F0 indicates mature production of Korean stop contrasts.

Statistical comparison with VOT and F0 in adult speakers' productions would be helpful in order to judge how close the toddlers are to reaching phonetic maturity. One crucial thing to note is that this is useful only when the phonetic values are comparable (Snedecor & Cochran, 1989). Using two-sample t-tests is appropriate for the comparison of VOT values in children's productions and adults' productions. Gender or age differences of VOT in consonant productions have barely been found (Koenig, 2001; Lee & Iverson, 2008; Nittrouer, 1993; Sweeting & Baken, 1982). On the other hand, F0 varies depending on physical factors, such as the size of vocal folds and sublingual pressure, so it is easily affected by gender or age. For this intrinsic and extrinsic variation, in this statistical method, this study ended up excluding the comparison of F0; only VOT in child speech was compared to VOT in the production of the adult control group using a two-sample t-test.[1]

3.5.2.2. The simple linear regression model

A simple regression model presupposes that there is a positive or negative correlation between a dependent variable and a predictor variable. That is, it assumes that there is a

---

[1] Since F0 is a so-called relative acoustic measure, I struggled to find the way to convert to an absolute measure. I thought of using z-scores rather than raw F0 values. However, using z-scores seems to be inappropriate in child speech since a mean value plays a significant role to transfer to a z-score. The mean F0 value in the child group cannot play a key role since the mean F0 actually means nothing in non-parametric distributional data. The child speaker group consists of various ages that critically affect the F0 values, so the mean of the total child participants cannot be used in transferring to z-scores. With this weakness, I decided not to use z-scores for this kind of analysis. Unfortunately, I was not able to find an appropriate absolute measure for F0, so this study was not able to conduct the comparison to the adult speakers' F0 values. Instead, this study is going to judge how successfully child speakers differentiate three stop groups in F0 through a simple linear regression model, which is discussed in 3.5.2.2.

linear relationship between a dependent variable and factors, and the relationship can be drawn as a line on a plot with the factor on the x-axis and a dependent variable on the y-axis. This property makes it useful to test the hypothesis that there is a clear relationship between a child's age and linguistic performance. To draw a clear line of correlation, I set up what the dependent and independent variables should be. The primary interest for this research would be the relationship between age and the mature development of the acoustic parameters, so the age of the child participants should be an independent variable and the actual VOT and F0 values are the dependent variables for this statistical model. The data was collected mainly because this study hypothesized that there should be a clear relationship between a speaker's age and their linguistic development, so this model primarily focuses on the effect of age in the given data. One thing to note is that mean VOT or F0 difference from the control group's productions were used as the dependent variables.

The main reason why this model had to use VOT difference or F0 difference rather than the raw measures as dependent variables is that we cannot judge the goodness or maturity of VOT and F0 values by themselves. The productions of the control group would be a good phonetic standard to determine the goodness of the children's sound tokens, so the VOT and F0 differences between adult and child productions were calculated and used as dependent variables. Each child speaker's average VOT and F0 values for each stop category were matched up with the average VOT and F0 values of all of the adult speakers for identical stop categories. With the absolute VOT and F0 differences between productions of adult and child speakers, this regression model suggests a simple clear relationship; the bigger the absolute differences, the more immature the production. With this simple regression model, the children's phonetic achievement is analyzed on the individual level. Since smaller VOT and F0 differences mean better results, it was expected that there would be a negative correlation between the values of the dependent variables and children's age.

3.5.2.3. The mixed-effects linear regression model

The previously introduced methods are basically designed for comparison with adult speakers' phonetic values, but the characteristics of the child data need to be analyzed in order to find any significant interaction between various predictors and the development of VOT and F0. The most appropriate way to determine the optimal statistical modeling starts from understanding the two representative characteristics of the obtained child data.

First, the data has unbalanced numbers of sound tokens depending on the speaker. This study did not control for age and gender of child participants, so the population in the participant group includes unbalanced numbers for gender and age. It would be practically impossible to interfere with child participants' performance during the experiments because it might affect their linguistic competence and performance. For the best result, which is the most natural and reliable outcome, the experimenter did not express any coercive behavior or gesture even when a child was very passive or distracted. Therefore, it should be naturally excused that the youngest child (age 1;8) provided a relatively small number of tokens and the oldest child (age 3;11) provided a large amount of tokens up to 54. In a case like this, which has unbalanced numbers of sound tokens, it is most effective to use average phonetic values, to represent a general tendency of child speech.

In spite of the benefit of using averages, this method also has its deficits. In the case of adult speech, mean phonetic values are commonly used, but in the case of child speech, each articulation can be extremely different depending on what context the target phoneme was pronounced in. The first impression I had in dealing with the acquired sound tokens was that the toddlers' pronunciations of the same stop phoneme are not consistent all the time. For example, '*childid_41*' showed some replacement for labial fortis stop; /p^h/ was substituted for /p'/ in many cases, but other fortis stops /t'/ and /k'/ were barely replaced with their counterparts from the different stop categories. Since the children's productions are not consistent, it is strongly suggested that every token has to be used in the child language study in order to judge how accurately and appropriately the children's phonetic implementations develop in detail. Every production by the same child speaker affects the judgment of their language competence, and it is not necessarily accurate to generalize to the linguistic ability of the same aged children.

Therefore, the child data has unbalanced numbers of tokens and average values cannot function as an average. Considering these two facts, statistical modeling can appropriately handle the data by labeling each repetition with two different levels: a production- and a speaker-level. Because the same child speaker repetitively produced the same phoneme, each articulation includes multi-level information; it simply depends on which speaker pronounced which phoneme. Through this multi-level statistical modeling, the data can be the most effectively and appropriately analyzed.

The primary benefit from using this mixed-effects model would be that it includes two error terms from the two different levels. That is, the data have non-observable individual-level and production-level differences and the two error terms automatically predict those differences. Another benefit would be that a mixed-effects model could explain the characteristics of the child data without any help of the adult productions. With a simple regression model, the degree of achievement in the acoustic implementation of the children could mainly be observed. In spite of the importance to judge their mature productions, any effect of correlations between predictor variables could not be observed in this way. Since the previous t-tests were also not able to help to illustrate a clear relationship between age and the development of VOT and F0 distinction, I still need to try to draw a clear line of age effects. With this strong benefit, the child data could be examined through a mixed-effects linear regression model.

3.6. Results

3.6.1. Adult productions

3.6.1.1. VOT-F0 distribution for Korean stop contrasts

VOT variations across the three different phonation types that were found in the production of the control group show that fortis stops tend to have a distinctive VOT range compared to the other two stop categories. As shown in Figure 3.3, all male and female adult speakers showed similar VOT variations depending on stop category in their productions. The general observation here is that aspirated stops have the longest VOT values, while fortis stops have the shortest VOT values and lenis stops have intermediate ones. The relatively shorter VOT in production of fortis stops recalls one of the recent reports that the most important phonetic definition and characteristic for fortis stops should be their shortest VOT among the stop types (Kong *et al.*, 2011). For lenis and aspirated stops, mostly overlapping VOT values have been found, which can be motivation for finding another acoustic parameter to distinguish them.

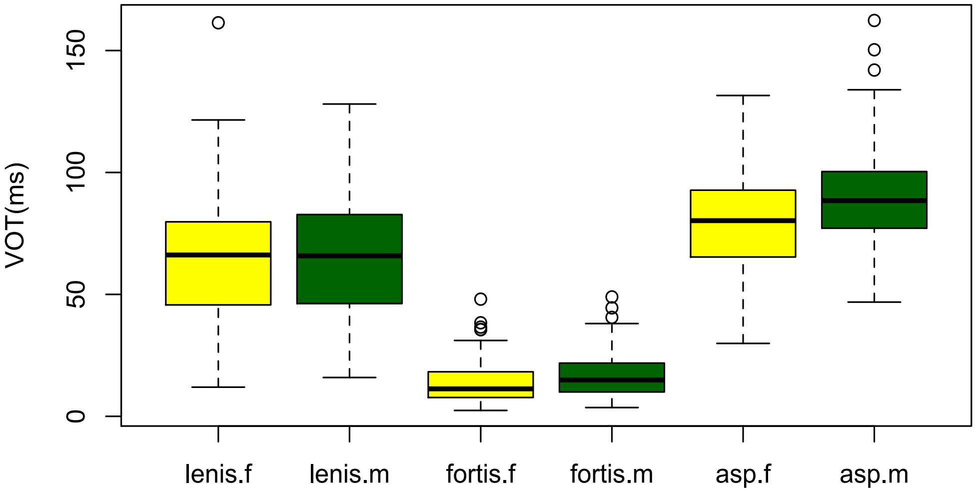

Figure 3.3. Boxplot for VOT in production of the control group. On the x-axis, ‘f’ indicates female speakers and ‘m’ indicates male speakers.

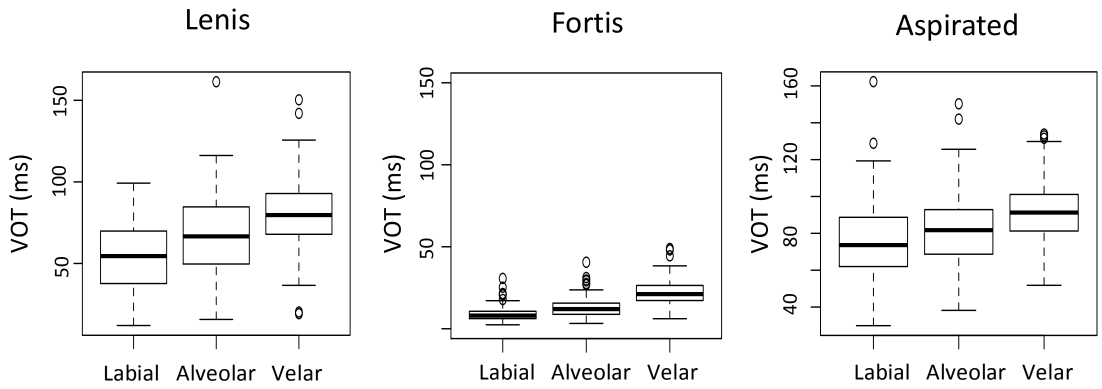

Figure 3.4. Boxplots for VOT in production of three stop groups with different POAs by control group.

VOT is a measure that directly relates to time and pressure, which means POA can affect VOT values. Figure 3.4 illustrates how different POAs can affect VOT values in adult speech. Considering acoustic and physiological aspects of VOT, as expected, within an identical stop category labial stops have the shortest VOT values, while velar stops have the longest values. Alveolar stops have intermediate values. This pattern is consistent with the cross-linguistic tendency that VOT becomes longer as the closure occurs at more posterior parts of the vocal tract (Maddieson, 1997). With this physiological reason, the effect of POA on the VOT values seems apparent here, but it should be noted that the VOT differences depending on POA are mostly overlapping in the same stop category and that different phonation types would make much bigger VOT differences overall. For example, it can be observed in Figure 3.4 that VOT values for velar fortis stops are much shorter than those of labial lenis stops. The effect of stop category on VOT must be much more determinant than that of POA.

F0 is a gender-dependent phonetic parameter since the pitch pattern is correlated with vocal cord size, so it is natural that the F0 values in the production of the control group are divided into two separate ranges. The production patterns are plotted separately according to gender in Figure 3.5. F0 ranges are clearly separated by the two different gender groups, but the overall patterns across the three different phonation types are almost identical. Lenis stops have the lowest F0 values, while aspirated stops have the highest values and fortis stops have intermediate values. At this point, it is not clear how significantly F0 values differ in each stop group, but it seems that the F0 difference between lenis and aspirated stops can provide a substantial phonetic difference between them, while their VOT values are similar.

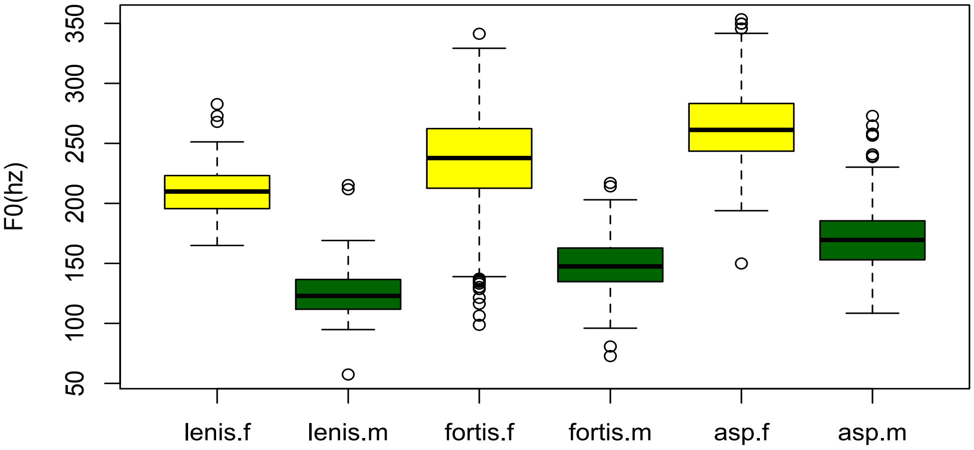

Figure 3.5. Boxplot for F0 in three stop categories by two gender groups of adult speakers. On the x-axis, ‘f’ indicates female speakers and ‘m’ indicates male speakers.

The following figure shows the effect of different POAs on the F0 ranges in stop productions. The two different gender groups showed similar F0 ranges across the three different stop categories depending on POA. In the same stop category, F0 values did not seem affected by POA. This is an observable acoustic difference from VOT in that F0 does not differ by place of closure in the vocal tract.

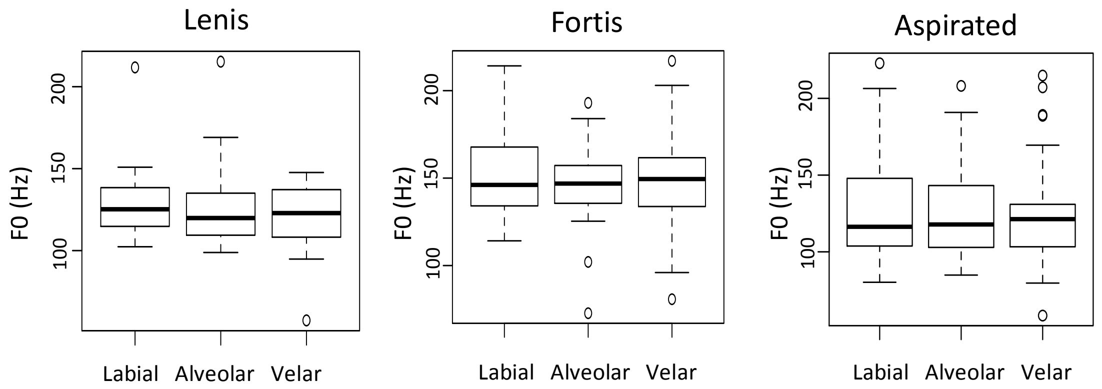

Figure 3.6. Boxplots for F0 in three stop groups with different POAs by male adult speakers.

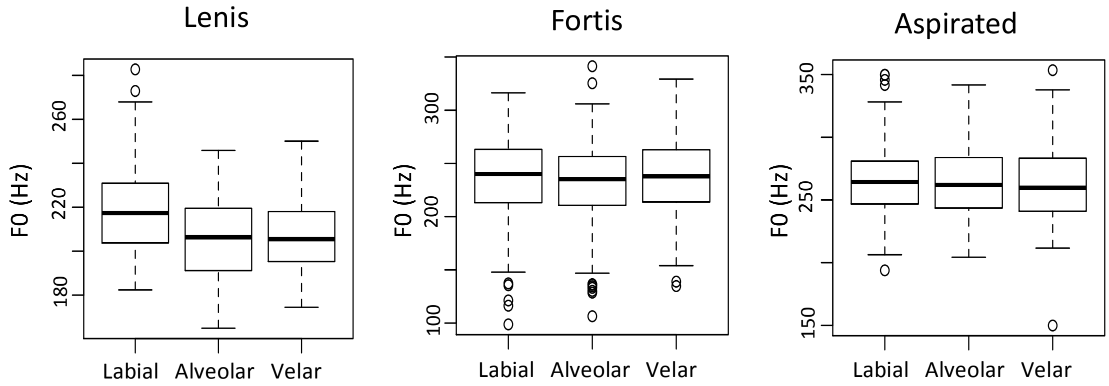

Figure 3.7. Boxplots for F0 in three stop groups with different POAs by female adult speakers.

To provide the standard dispersion patterns for the three stop categories in the two representative acoustic dimensions of VOT and F0, male and female adult groups' sound tokens are plotted. In general, the two gender groups' F0 ranges are clearly different, but the two groups showed similar relative phonetic differences between the stop contrasts.

Fortis stops have observable distinctive VOT values compared to lenis and aspirated stops, which implies that the role of VOT is determinant with no major dependency on F0 to characterize fortis stops phonetically in production. In the case of the productions of the male adult speakers, VOT values for lenis and aspirated stops are mostly dispersed over a large range from 10 ms to 120 ms, while F0 values in those two stop groups are relatively concentrated within a compact area. In particular, F0 values in lenis stops are even more closely gathered than those in aspirated stops.

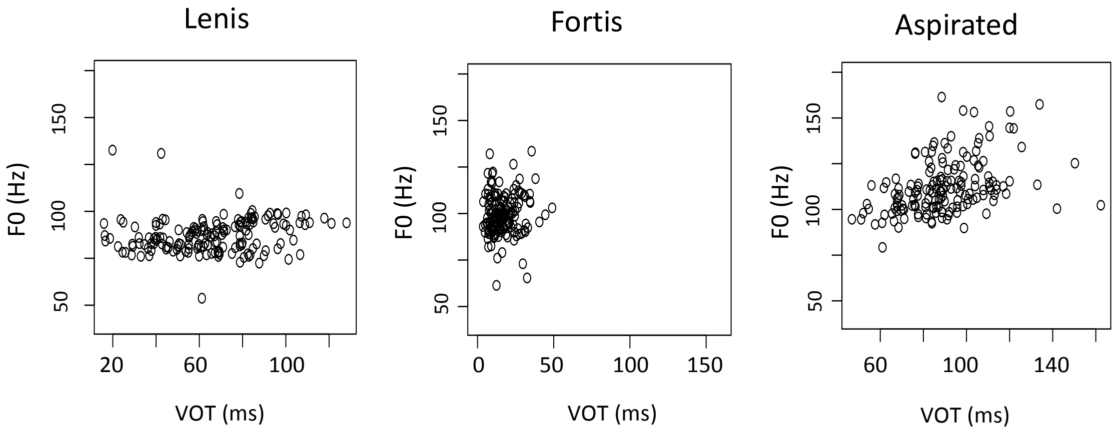

Figure 3.8. Scatterplots for every token by male adult speakers with VOT (x-axis) and F0 (y-axis).

In productions of the female adult speakers, the dispersion patterns are almost identical to the ones in the male speakers' production. The most distinctive difference is the concentrated VOT values in production of fortis stops. VOT values are all gathered in a small area of 0 to 50 ms. Even though we see some VOT overlaps across the three different phonation types, it can be confirmed that fortis stops must be fully distinguished phonetically by their shortest VOT range among the stop types.

On the other hand, most of the F0 values in lenis stops are gathered around 200 Hz, while VOT values are dispersed with the maximum of 120 ms-difference. Lenis and aspirated stops actually share an almost identical VOT range, so in the dimension of VOT, it hardly seems that the two stop manners would be phonetically differentiated. The crucial phonetic difference between lenis and aspirated stops should be their pitch patterns. The F0 difference between aspirated and lenis stops is up to 60 Hz for the adult speakers, which confirms that F0 is an important phonetic cue to differentiating the contrast between lenis and aspirated stops.

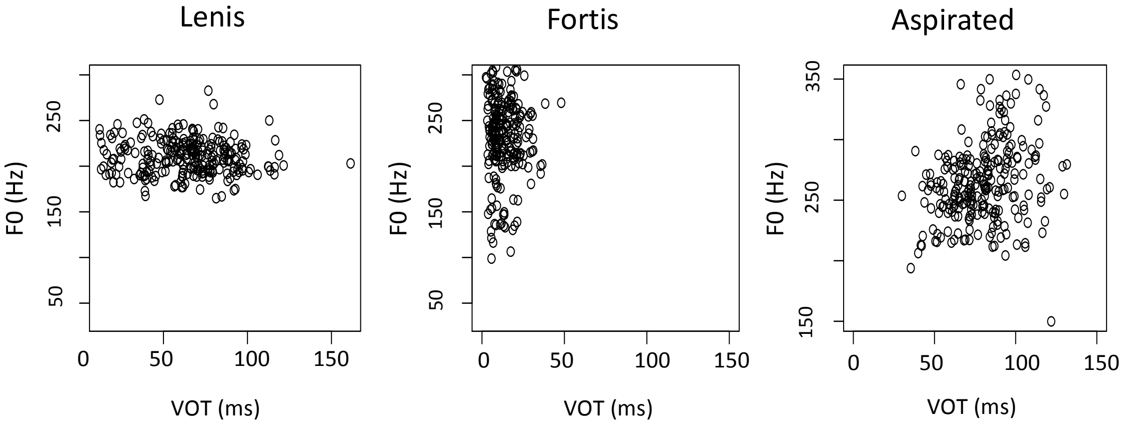

Figure 3.9. Scatterplots for every token by female adult speakers with VOT (x-axis) and F0 (y-axis).

The mean-difference by speaker of stop contrasts in the dimension of VOT and F0 is well illustrated in Figure 3.10. Considering the fact that the adult speakers were in their late 20s or early 30s at the time of the recording, and therefore were born in the 1980's, their production patterns can represent the recent trend of Seoul Korean. However, it should be noted that their productions were recorded from the experimental setting with using a carrier sentence. This data obtaining process might affect the acoustic properties in their productions. Target words in a carrier sentence easily get exaggerated phonetic properties to some extent and might be differentiated in a traditional distinction way compared to their everyday natural speech. Still, their productions show the phonetic distance in VOT between fortis and the other two non-fortis stop categories, while subtle VOT differences can mostly be found in lenis and aspirated stops. The phonetic distance in F0 for those two stop categories functions as the offset to this VOT overlap between them. This pattern has been consistently observed and can be evidence of a tonogenetic sound change in Seoul Korean.

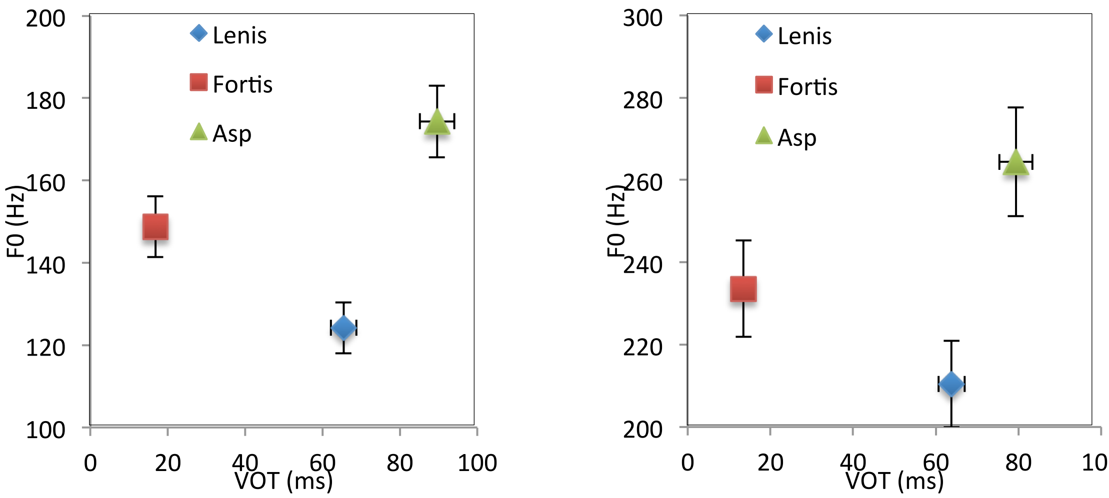

Figure 3.10. Male (left panel) and female (right panel) speakers' difference in mean VOT-F0 by phonation type. Error bars represent 95% confidence interval.

3.6.1.2. Statistical analysis

In this section, the observed phonetic differences are statistically analyzed. It is crucial to see how significantly those phonetic differences affect the formation of the acoustic characteristics of each phonation type. The primary purposes of modeling statistical methods are to answer whether the shortest VOT values for fortis stops would be a uniquely determinant phonetic definition and whether the F0 values at the following vowel onset alone could produce a significant phonetic difference between lenis and aspirated stops such that F0 could exclusively function to distinguish them.

To answer these questions, two different statistical models were designed and run: a two-sample t-test and a mixed-effects linear model. The two-sample t-test compares the VOT values or F0 values of pairs of lenis and fortis stops, lenis and aspirated stops, and fortis and aspirated stops, and allows us to investigate how significantly VOT values or F0 values in different stop groups are different. A mixed-effects linear model shows the effects of each predictor variable such as gender or stop category on VOT or F0 variations. The same speaker pronounced the same word three times repeatedly, and this model uses all the acquired tokens, so two-level treatment was required for a more accurate analysis. A linear mixed-effects model can handle this kind of multi-level information, so it is useful for the analysis.

The results of the two-sample t-tests in the tables below are unexpected.

Table 3.3. The output of the two-sample t-tests for VOT.

Two-way t-tests	*p*-value
VOT comparison in fortis and lenis stops	$p < 0.05$
VOT comparison in fortis and aspirated stops	$p < 0.05$
VOT comparison in aspirated and lenis stops	$p < 0.05$

Table 3.4. The output of the two-sample t-tests for F0.

Two-way t-tests	*p*-value
F0 comparison in fortis and lenis stops	$p < 0.05$
F0 comparison in fortis and aspirated stops	$p < 0.05$
F0 comparison in aspirated and lenis stops	$p < 0.05$

In both cases, for male and female adult speakers, all the acquired *p*-values were similarly small ($p < 0.05$), indicating that VOT or F0 values in different stop groups are significantly different. This result can be interpreted to mean that adult speakers' VOT and F0 in each stop category are so standardized with no outliers that the observed phonetic differences are considered significant. This is an interesting finding since it has been discussed by many researchers that Seoul Korean has recently been undergoing a tonogenetic change, and therefore a phonetic distinction in the dimension of VOT is lost, especially between lenis and aspirated stops (Kang, 2014; Silva, 2006; Wright, 2007).[2]

The second statistical method is a multi-level model. I constructed a mixed-effects linear regression model implementing the *lme4* package (Bates *et al.*, 2015) in R (R Development Core Team, 2011). VOT values were dealt with without considering gender difference, but F0 values were applied separately according to speakers' gender.

---

[2] As mentioned in Section 3.6.1.1, adult speakers' productions were recorded in an experimental setting, which can produce exaggerated phonetic differences between contrasts. In their natural speech, it is possible that VOT differences would not be significant as found in the recorded production of target words in a carrier sentence.

In order to see the independent effects of predictor variables, the equation in (1) was designed. *Lenis stops* and *Female* were the reference categories, for *Stop Category* and *Gender* respectively in the analysis. As represented in (1), VOT is the independent variable, the fixed effects predictors are *Gender* and *Stop Categories*, and by-speaker adjustment to intercept is a random effect.

(1) $VOT_{ij} = \beta_0 + \beta_1 \times Male_i + \beta_2 \times Fortis_{ij} + \beta_3 \times Aspirated_{ij} + \eta_i + \varepsilon_{ij}$

($_i$ = speaker level, $_j$ = production level)

Table 3.5. The output of the mixed-effects linear model for effects on VOT in adult production.

Random effects:

Groups	Name	Variance	Std. Dev.
id	(Intercept)	81.85	9.047
Residual		283.38	16.834

Number of obs: 1235, groups: id, 24

Fixed effects:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	62.476	2.585	24.172	< 2e-16 ***
Fortis	−49.765	1.174	−42.399	< 2e-16 ***
Asp	18.971	1.172	16.184	< 2e-16 ***
Male	5.07	3.988	1.271	0.217

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The gender difference does not make any significant phonetic difference in VOT ($p > 0.05$). However, each phonation type shows significant VOT differences between the different stop categories. Fortis stops tend to have significantly shorter VOT (*coef.*= −49.765, $p < 0.001$), while aspirated stops tend to have significantly longer VOT (*coef.* = 18.971, $p < 0.001$). This finding is consistent with the results of the two-way t-test that showed significant VOT differences between Korean stops.

Equation (2) includes an additional interaction term '*Stop Category* × *POA*' in order to show the effect of an interaction between each stop category and POA on VOT. Here, *Lenis stops* and *Labial* were the reference categories.

$$
\begin{aligned}
(2)\ \mathit{VOT}_{ij} = {} & \beta_0 + \beta_1\times \mathit{Fortis}_{ij} + \beta_2\times \mathit{Aspirated}_{ij} + \beta_3\times \mathit{Alveolar}_{ij} + \beta_4\times \mathit{Velar}_{ij} \\
& + \beta_5\times \mathit{Fortis}_{ij}\times \mathit{Aveolar}_{ij} + \beta_6\times \mathit{Aspirated}_{ij}\times \mathit{Alveolar}_{ij} + \beta_7\times \mathit{Fortis}_{ij}\times \mathit{Velar}_{ij} \\
& + \beta_8\times \mathit{Aspirated}_{ij}\times \mathit{Velar}_{ij} + \eta_i + \varepsilon_{ij}
\end{aligned}
$$

Table 3.6. The output of the mixed effects linear model for the effect of stop category and POA on VOT in adult production.

Random effects:

Groups	Name	Variance	Std. Dev.
id	(Intercept)	85.9	9.268
Residual		228.8	15.128

Number of obs: 1235, groups: id, 24

Fixed effects:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	52.434	2.322	22.579	< 2e-16 ***
Fortis	−43.455	1.821	−23.862	< 2e-16 ***
Asp	23.28	1.821	12.783	< 2e-16 ***
Alveolar	13.809	1.821	7.583	6.75e-14 ***
Velar	22.35	1.825	12.25	< 2e-16 ***
Fortis:Alveolar	−9.962	2.583	−3.587	0.000121 ***
Asp:Alveolar	−7.282	2.578	−2.825	0.004806 **
Fortis:Velar	−9.031	2.58	−3.5	0.000482 ***
Asp:Velar	−5.658	2.58	−2.193	0.028498 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The coefficients of all cases show that the variables have significant effects on VOT in adult speakers' productions (*p*-values < 0.05). This means that the interaction between each stop category and different POAs results in significant VOT differences. This implication was previously illustrated in the boxplots for adult stop productions (Figures

3.6 and 3.7). In order to determine whether a fixed effect of the interaction between POA and stop categories is significant, an *ANOVA* was conducted. It confirmed that the correlation between POA and each stop category is significant ($X^2 = 264.94$, $df = 6$, $p < 0.001$). In addition, the *lsmeans* function with "Tukey" from the *lmerTest* package (Kuznetsova *et al.*, 2013) was implemented for a multiple comparison of the fixed effects and their interactions. These pairwise tests reported that the VOT differences between labial and alveolar fortis stops were not significant ($p = 0.4733$) but for the other two stop categories, lenis and aspirated, POA effects on VOT were all significant (p-values < 0.05). The biggest mean VOT difference was found between labial and velar lenis stops (= 22.35 ms).

The next model deals with the correlation between F0 and phonation type in the production of the adult speakers. F0 ranges differ extremely depending on gender, so the linear regression model was applied separately by gender. The formula in (3) was constructed to determine the effect of each stop category on F0 values. *Lenis stops* were the reference category.

(3) $F0_{ij} = \beta_0 + \beta_1 \times Fortis_{ij} + \beta_2 \times Aspirated_{ij} + \eta_i + \varepsilon_{ij}$

Table 3.7. The output of the mixed effects linear model based on (3) for male adult speakers.

Random effects:

Groups	Name	Variance	Std. Dev.
id	(Intercept)	237	15.4
Residual		351	18.73

Number of obs: 486, groups: id, 10

Fixed effects:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	124.226	5.339	23.27	2.91e-09 ***
Fortis	24.543	2.082	11.79	< 2e-16 ***
Aspirated	50.117	2.082	24.08	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Table 3.8. The output of the mixed effects linear model based on (3) for female adult speakers.

Random effects:

Groups	Name	Variance	Std. Dev.
id	(Intercept)	395	19.88
Residual		729.3	27.01

Number of obs: 749, groups: id, 14

Fixed effects:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	210.483	5.579	37.73	4.44e-16 ***
Fortis	23.159	2.418	9.577	< 2e-16 ***
Aspirated	53.993	2.413	22.375	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

As confirmed statistically, the two different gender groups showed similar production patterns. The mean F0 values in production of lenis stops in both gender groups have some F0 difference (86.257 Hz = 210.483 – 124.226), but the relative F0 differences from fortis or aspirated stops are almost the same (24.543 Hz vs. 23.159 Hz; 50.117 Hz vs. 53.993 Hz). The estimated F0 differences between different stop groups are significantly large ($p < 0.001$), which is consistent with the results from the two-way t-tests. The most important thing here is that F0 differences between lenis and aspirated stops differed the most, with a maximum of 50 Hz.

Statistical analysis showed that VOT is the shortest for fortis stops and is significantly influenced by interaction with different POAs, as long as they are in the same stop category. F0 values also differ depending on which stop category precedes the vowel, and the difference in F0 between lenis and aspirated stops is greatest for both gender groups, around 50 Hz in standard adult speakers' productions.

3.6.2. Child productions

A total of 58 child subjects participated in the production experiment, and VOT and F0 values from their productions were calculated for analysis. This process allowed us to determine if there exists an ordering of mastery between the two acoustic parameters, VOT and F0, and if there is a significant correlation between age and the development of VOT or F0. Through this analysis, it can be observed how close or how far the children's stop distinctions may from the adult speakers' articulatory stop distinctions. If their distinguishing methods are different from those of the adult speakers, then other methods of phonetic discrimination can be studied. Here, through the analysis, it can be illustrated how VOT and F0 operate in distinguishing Korean stop manners and how the developmental patterns of the two acoustic parameters may differ according to age.

3.6.2.1. Phonetic difference in the dimensions of VOT and F0

Overall, child participants showed diverse patterns in their VOT ranges over the Korean stop categories compared to the adult speakers. This freely dispersed VOT pattern can be found even within the same child speaker's productions. Phonetic inconsistency in the production of the same phoneme is also observed depending on the context in which the target phoneme is included. This inconsistent and unstable articulation indicates that the children are still learning how to produce phonetic differences for stop contrasts in terms of VOT. In spite of this general tendency, successful discrimination between the three different phonation types was found in the case of relatively older children's productions. VOT is a representatively universal phonetic feature of phonological stop contrast in many languages (e.g., Cho & Ladefoged, 1999; Lisker & Abramson, 1964), so it was fully expected that VOT would be easy to acquire at an early age. The general articulation pattern for each stop category is shown in Figure 3.11. It was plotted with all acquired production tokens from every child participant.

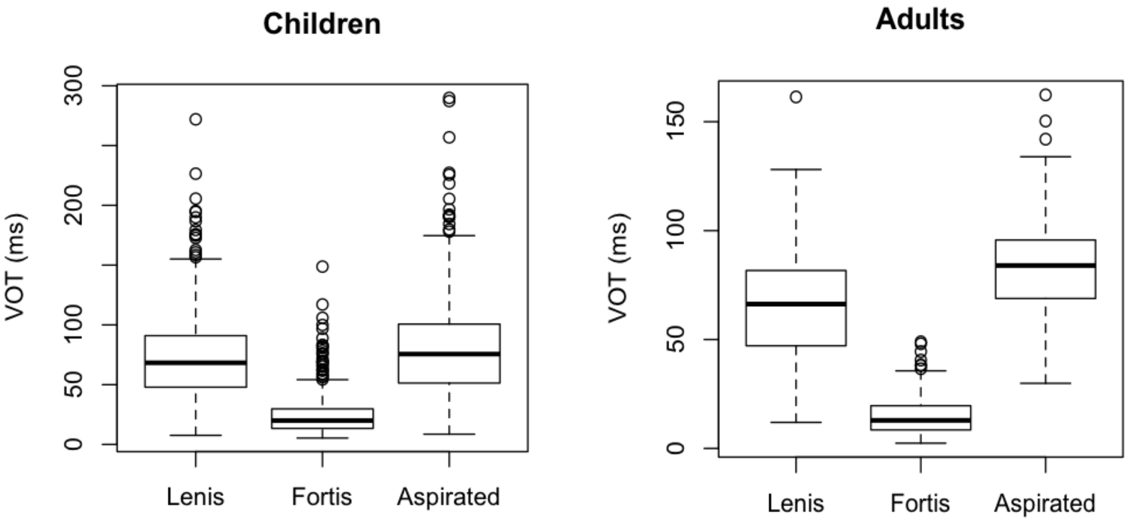

Figure 3.11. Boxplots for VOT values in the production of all child (left panel) and adult (right panel) participants.

Because of extreme inconsistency in VOT even within a given child speaker's productions, the VOT data for the three stop categories includes a lot of outliers. In spite of the large number of outliers, it can still be found that the VOT values for fortis stops are clearly shorter than those for lenis or aspirated stops. Lenis and aspirated stops share almost identical ranges of VOT, but in the case of aspirated stops, a slightly wider range of VOT values is found. The figure shows that VOT does function in distinguishing for fortis stops in the children's productions. The notable difference from the adult speakers' VOT variation is that the VOT ranges overlap for lenis and aspirated stops, and it is also notable that the VOT difference between fortis and non-fortis stops tends to be much smaller than that in adult speakers' productions.

The following figures provide another boxplot for VOT in three stop categories produced by the young children. This was plotted only with the mean VOT value of each child participant. The main reason for this is to minimize the visual effects of the outliers as in Figure 3.11, and to show an average pattern with the mean differences between the contrasts. The average pattern confirms that lenis and aspirated stops have overlapping VOT values, so it seems difficult to distinguish them solely by VOT. On the other hand, fortis stops have a distinctive VOT range, which clearly distinguishes them from non-fortis stop categories. Considering the fact that the speakers are toddlers who are learning language, it is somewhat surprising that they are already able to differentiate fortis stops using VOT difference.

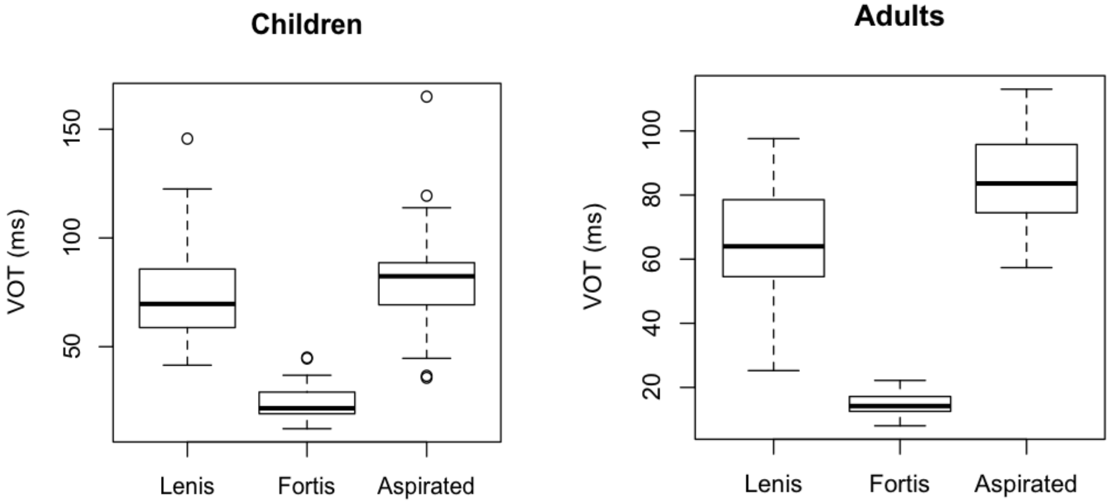

Figure 3.12. Boxplots for by-speaker mean VOT in children's (left panel) and adults' (right panel) speech.

The next issue is the relation between age and VOT values in production. In order to see how age and VOT variation in each stop category are related, scatterplots with age as x-axis and VOT as y-axis are illustrated as follows. The mean VOT value for each child speaker was calculated and used, mainly because this would make it easier to see how differently children produce the same stop group. Another reason is that every speaker provided different numbers of sound tokens so using every sound token was avoided.

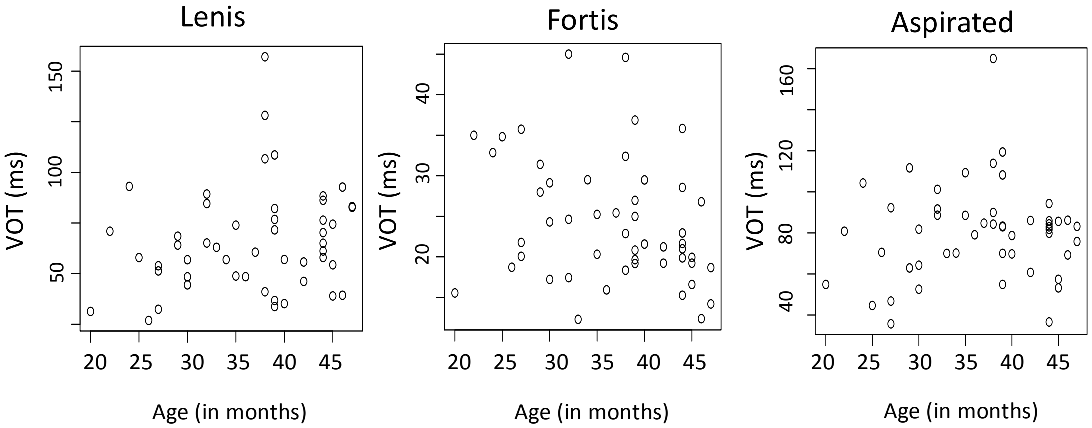

Figure 3.13. Scatterplots for mean VOT of each child speaker in relation of age to VOT.

It is shown that in the production of lenis stops, VOT values are widely dispersed from 30ms to 150ms, and this tendency is easily found at every age. The relatively older children aged 40 to 45 months also showed a 40 ms-difference in the production of lenis stops depending on the speaker. However, VOT values in fortis stops are within a relatively small range (from 5 ms to 45 ms), and relatively older children produced fortis stops with much shorter VOT values, ranging from 10 ms to 35 ms but mostly gathered around 20 ms. It is also notable that from the younger aged to older aged children, their VOT values are similarly estimated; this phenomenon represents that the shortest VOT values can be a crucial phonetic definition for fortis stops and that this can be achieved at very early ages.

Similarly to the productions of lenis stops, VOT values in aspirated stops ranged widely from 20 ms to 170 ms. At a glance, in the productions of lenis and aspirated stops, a close relation between VOT and age would not be easily predicted since older children also showed widely dispersed VOT values in these two non-fortis stop categories. However, this does not necessarily mean that the acquisition of VOT did not occur. Since VOT would not directly affect lenis-aspirated distinction (e.g., Kang, 2014 for findings in adult productions), it seems appropriate that the acquisition of VOT should be dealt with in the production of fortis stops. As a result, the acquisition of VOT took place in early age and the role of VOT is determinant only for fortis stops.

As in the adult productions, it has been questioned if POA affects the VOT variations in the production of the child participants. Generally, it is a well-known fact that velar stops tend to have longer VOT than labial or alveolar counterparts, which was shown in the case of the control group. The POA effects on VOT can be another criterion to judge how successful children's VOT control is. However, this tendency has not been easily found in the case of child speakers here. There are some VOT differences observed, but it is unclear whether such differences depend on truly different POAs, since phoneme substitution is a common sound error found in toddlers' productions (Van & Erickson, 1996). The unique pattern in Figure 3.14 is that the labial lenis stop /p/ and the labial fortis stop /p'/ have a relatively large number of outliers, which means VOT values in children's productions of labial stops are widely dispersed. Moreover, immature VOT patterns are found in aspirated stops. In the production of aspirated stops, labial stop /p^h/ has longer VOT values than /t^h/ or /k^h/. Children's diverse VOT variations include the fact that the target phoneme might not be pronounced in a standard way, since the effect of POA on VOT is a phenomenon that shows the intrinsic characteristics of VOT. Even though the toddlers tried to produce a VOT difference for fortis stops in some way, their stop productions are still immature.

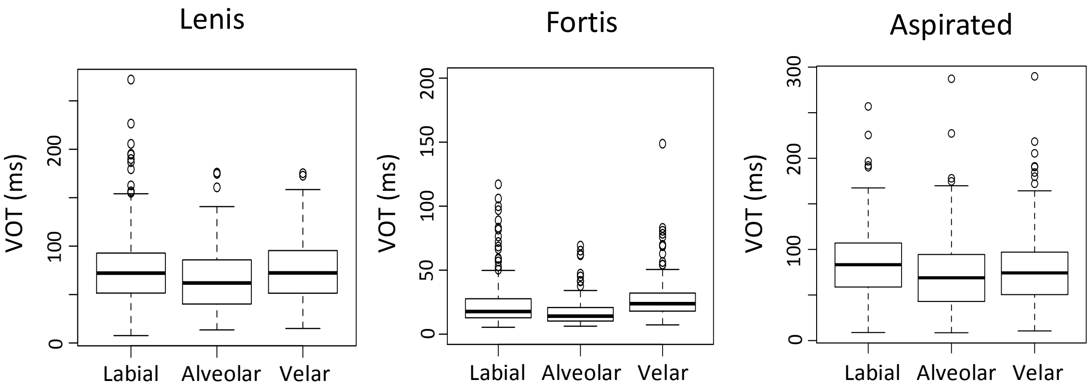

Figure 3.14. Boxplots for the POA effects on VOT in child speakers' productions of three stop categories.

The second important acoustic parameter in distinguishing Korean three-way stop contrast is F0. As shown in the previous explanations, VOT crucially operates in defining fortis stops, while VOT values in lenis and aspirated stops are similar. To make a solid phonetic distinction between lenis and aspirated stops, F0 is crucial. F0 values at a following vowel onset differ depending on what phonation type precedes the vowel. Focusing on the property of F0 perturbation, the following figure illustrates the F0 ranges in distinguishing Korean stop categories by the children. The F0 values are quite widely dispersed from 100 Hz to 600 Hz for all three stop categories, but the median in each category shows some F0 differences: Aspirated stops have the highest F0 average, while lenis stops have the lowest F0 average. This general pattern is consistently found in the productions of the control group, even though the degree of F0 difference is more pronounced in the case of adult speech (Figure 3.5).

One clear observation in toddler production is that fortis and aspirated stops have nearly overlapping F0 values. F0 differences between fortis and aspirated stops in the productions of the control group were much more obvious, since their F0 values for those stop types were considered significantly different ($p < 0.05$). At this point, it can be said that the child speakers did not distinguish F0 for fortis and aspirated stops as the adult speakers did, though such a large F0 difference between fortis and aspirated stops for the adults seems somewhat redundant since fortis stops have considerably shorter VOT values than aspirated stops.

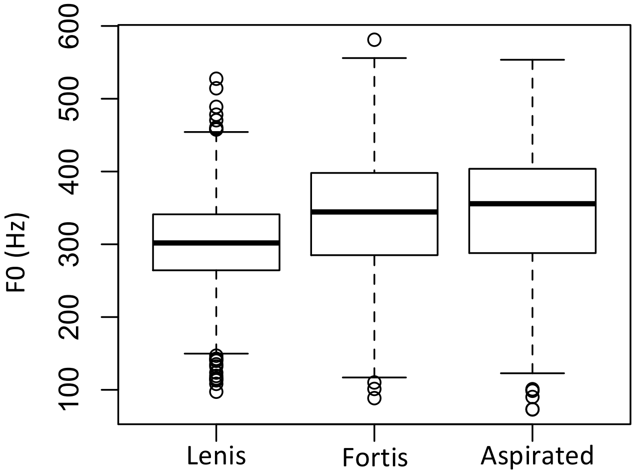

Figure 3.15. Boxplot for F0 in the production of three stop categories by child speakers.

The F0 variation in production in relation to the child age is illustrated in Figure 3.16. No specific phonation type shows any concentrated region for F0 values at vowel onset. In general, all the sound tokens were widely dispersed in all three cases. The dispersion is found even in older children's productions, so it is hard to see a correlation between age and F0 patterns in stop production. That is, it cannot be determined that a specific F0 region phonetically defines a particular phonation type or that the older children's performance would be better in discriminating Korean stop categories than what the younger children did. At this point, the relation between F0 and a certain stop category also cannot be predicted through these scatterplots. Compared to VOT variations, it seems to be much more difficult for young children to acquire and use F0 as an acoustic tool to discriminate Korean stop contrasts.

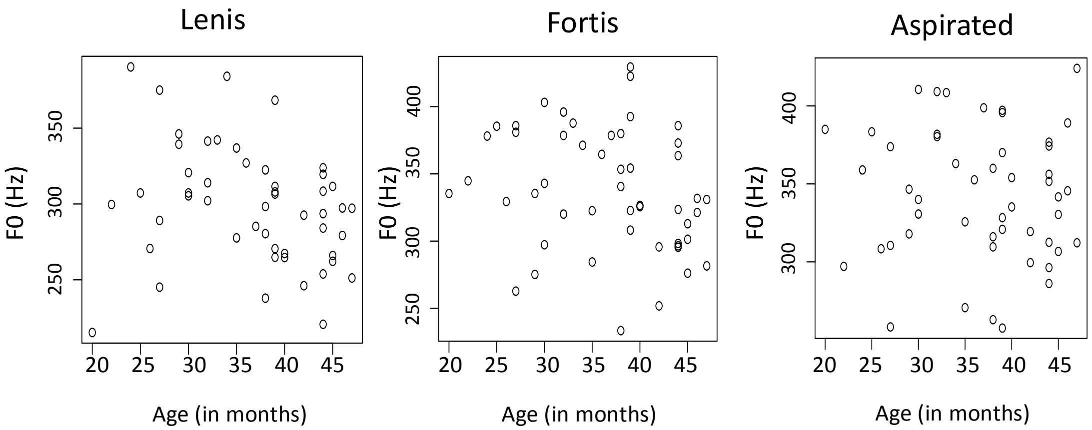

Figure 3.16. Scatterplots for mean F0 of each child speaker in relation to their age.

In spite of the fact that F0 distinction is not clear in the child speakers' productions, it is still questionable whether POA affects F0 variation in any way, as in the case of VOT. The following boxplots were drawn to investigate the possible POA effects on F0 in stop production.

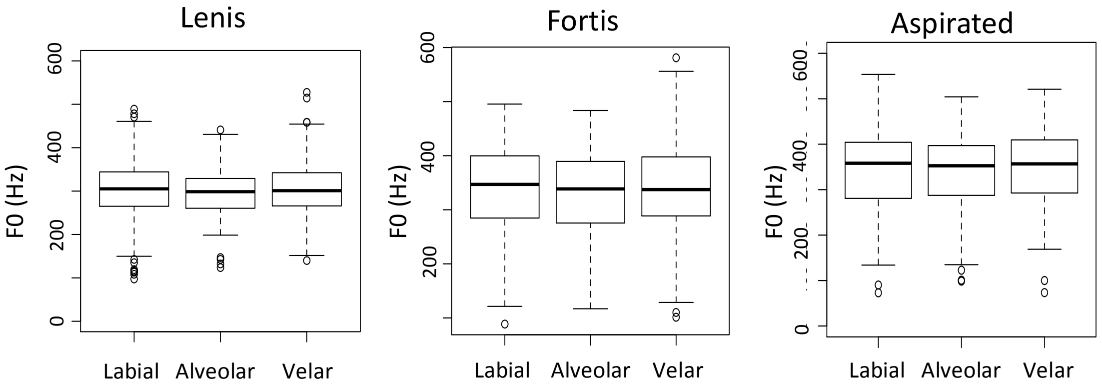

Figure 3.17. Boxplots for F0 in three stop groups with POA effects.

As shown in Figure 3.17, it is obvious that POA does not influence F0 differences within the same stop category. In the case of lenis stops, there are many outliers, especially for labial /p/, but in general, no POA effects are observed since the F0 variations are nearly the same. For the aspirated stops, the three different POAs share overlapping F0 ranges, and most of the sound tokens have F0 values from around 300 Hz to 400 Hz.

The following figures illustrate how in child production three stop categories can differ in the dimensions of VOT and F0. For plotting, VOT and F0 values were calculated from each child's average articulation. These scatterplots indicate that the overall production patterns are two-fold: fortis stops have clearly shorter VOT values, and for all three stop categories, F0 values are widely ranged with overlaps.

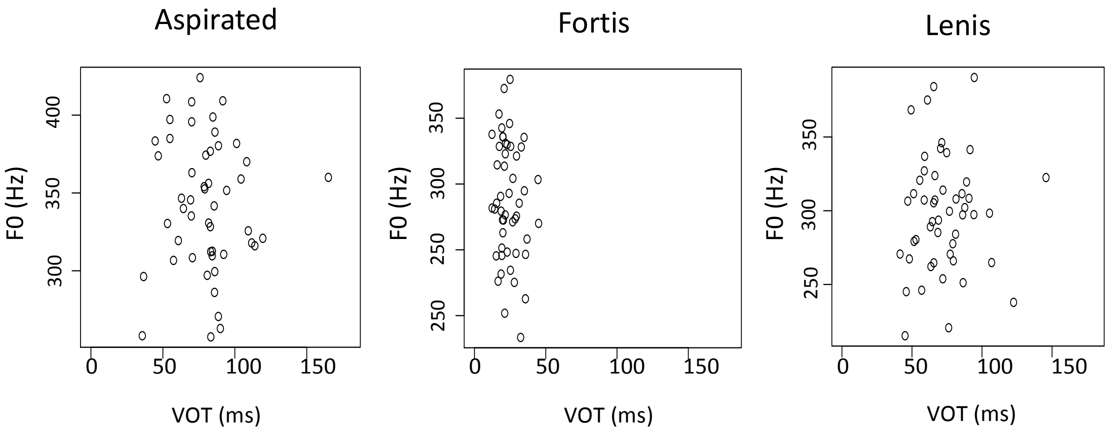

Figure 3.18. By-speaker mean stop production of the children in the dimensions of VOT (x-axis) and F0 (y-axis).

As expected, in fortis stops, even though VOT values are concentrated within a small area around 25 ms, F0 values are widely dispersed from 200 Hz to 400 Hz. This implies a close relationship between VOT and fortis stops. Productions in aspirated and lenis stops have much longer VOT values (about 40 ms longer than fortis stops), but a phonetic distinction between the non-fortis stop categories with VOT is not found, since their VOT values are nearly overlapping. This observation indicates that their phonetic distinction needs more substantial phonetic difference in the F0 dimension. Through the plots, we can see that the actual F0 difference between aspirated and lenis stops produced by the child participants exists to a small extent. As we recall, the estimated F0 difference between these two stop categories in the control group's production is around 60 Hz, but the observed F0 difference in child speech is not very substantial. This production pattern means that the child speakers seem to be aware of phonetic differences between their native stop contrasts.

The following plots in Figure 3.19 show how every token of the child speakers can be plotted in the same dimensional frame. Using all the tokens in plotting, the result seems to be distorted because of the unbalanced numbers of sound tokens per speaker. However, at the same time, it is important to go over every production pattern since toddlers' productions vary by many factors including the context. The density in these plots represents the same production pattern in Figure 3.18. While productions of fortis stops have VOT values around 10-20 ms, the VOT distributions of lenis and aspirated stops resemble each other; the most crowded region is from 20 ms to 100 ms. With respect to F0, fortis stops have the most dispersed F0 values, since most of the tokens are ranged from 200 Hz to 500 Hz. The sound tokens of aspirated stops seem to be a bit more dispersed than those of lenis stops in terms of VOT, but their dispersion patterns are similar in that their F0 values are ranged from 100 Hz to 500 Hz. Meanwhiles in the case of lenis stops, most tokens are gathered around 300 Hz.

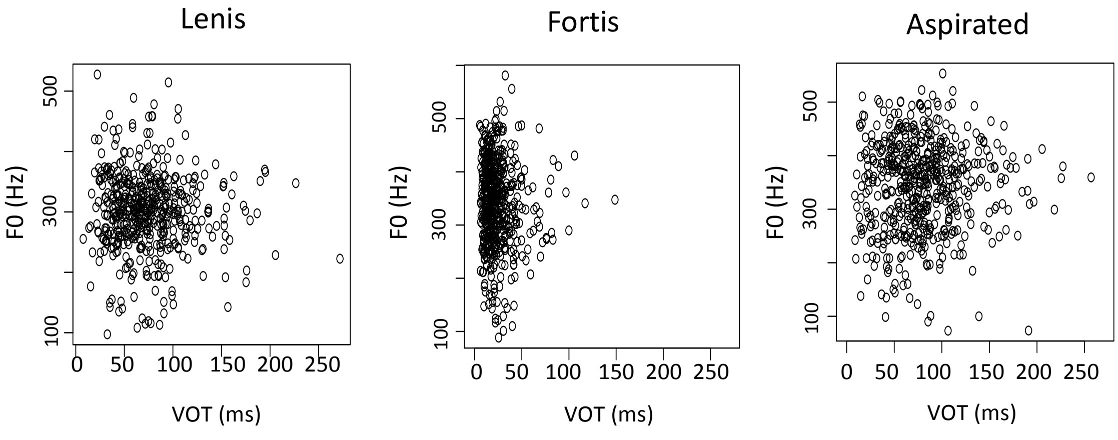

Figure 3.19. Production patterns of child speakers in the dimension of VOT (x-axis) and F0 (y-axis). All sound tokens were plotted.

The different figures above convey basically the same idea. The children's articulations of fortis stops can be characterized as having the shortest VOT, but the wide dispersion pattern of F0 indicates that the young children do not yet utilize F0 as a crucial acoustic tool for the stop distinction. Therefore, this production pattern amplifies the fact that VOT is acquired earlier than F0; this mastery ordering also affects the acquisition of a specific phoneme which is phonetically more directly related to VOT, while articulatory confusion between the other stop contrasts, lenis and aspirated stops, is found. VOT has been quite successfully shown as an acoustic parameter to distinguish fortis stops even for young children.

3.6.2.2. Statistical analysis for child production

3.6.2.2.1. T-test

With respect to VOT, a two-sample t-test was conducted in R (R Development Core Team, 2011) to determine whether each child speaker would be able to differentiate three different stop categories in their speech. The two-sample t-tests were performed three times per child. VOT values in productions of lenis, fortis, and aspirated stops by one child speaker were calculated to compare in three different pairs (lenis-fortis stops, lenis-aspirated stops, and fortis-aspirated stops). For this comparison, all the recorded sound tokens of all child speakers were used. The results are shown in the table below. In Table 3.9, the obtained *p*-values are indicated with a significance level.

In the child speakers' productions, VOT differences between fortis and lenis or aspirated stops are clearly shown across all ages. Except for the youngest child (*childid_48*, aged 20 months), the child participants in general could produce some phonetic difference in VOT for fortis stops, since VOT differences between fortis and non-fortis stops are generally significant according to the test results. In particular, from the age of 32 months, the older subjects had significant VOT differences between fortis and aspirated or lenis (p-values < 0.05). On the other hand, the younger children showed inconsistent VOT differentiation for fortis stops. A notable result is that there are no significant VOT differences between lenis and aspirated stops. This tendency is consistently found across all ages. As the control group's productions suggested, the overlapping VOT values of lenis and aspirated stops mean that they have to be distinguished in another dimension, F0. Through this statistical method, the degree of VOT difference in each child speaker's productions can be studied, but it is still unclear whether the estimated VOT differences resemble mature production patterns by adult speakers. To determine this, the control group provides the phonetic standards for VOT ranges for three stop categories, and another t-test was conducted for comparison with the child data.

Table 3.9. Results of two-sample t-tests of VOT in child production data.

Age (in months)	childid	Lenis-Fortis	Lenis-Aspiated	Fortis-Aspirated
20	48	0.07254 .	0.6723	0.1582
22	49	0.0003516***	0.7286	0.02938*
24	46	0.0008161***	0.4455	0.0002354***
24	58	0.6663	0.1416	0.1922
26	26	0.02739*	0.5383	0.5577
26	40	0.002152**	0.03018*	0.00104**
27	24	0.1345	0.3367	0.7867
27	39	0.01433*	0.4275	0.04398*
27	42	0.0001221***	0.8326	0.02575*
29	44	9.136e-05***	0.2147	0.001171**
29	45	0.1039	0.2612	0.1192
30	23	2.307e-06***	0.421	2.444e-05***
30	38	0.007447**	0.2249	0.01332*
30	41	0.07564.	0.923	0.0357*
31	57	0.0001982 ***	0.5132	0.137
32	25	1.481e-05***	0.04744*	1.812e-06***
32	43	0.002593**	0.9863	0.01057*
32	47	0.001822**	0.5374	4.402e-05***
32	53	0.012 *	0.05802 .	1.468e-07 ***
33	22	6.501e-07***	0.6516	9.405e-07***
33	55	0.03634 *	0.08251 .	0.006804 **
34	50	0.0006711***	0.7446	0.002042**
34	51	5.351e-07 ***	0.5638	0.005233 **
34	54	3.276e-05 ***	0.9287	0.005395 **
35	4	2.756e-05***	0.03915*	9.978e-07***
35	13	0.003304**	0.1758	0.004541**
36	16	2.965e-05***	0.05964.	4.36e-09***
37	33	0.001973**	0.3585	0.005984**
38	1	2.465e-05 ***	0.29	0.0005113 ***
38	17	2.398e-05***	0.1849	0.004064**
38	35	0.5066	0.01598*	0.01193*
38	36	0.0001806***	0.2916	0.0004572***
39	2	7.764e-06***	0.7707	4.042e-07***
39	6	3.202e-07***	0.1409	5.658e-07***
39	10	1.846e-09***	0.1825	0.0001332***
39	31	0.0004456***	0.4799	0.000336***
39	32	1,335e-05***	0.8447	3.238e-06***
39	37	0.1338	0.06476.	0.007481**
40	14	0.001678**	0.1773	0.0003155***
40	29	0.03877*	0.09991.	4.499e-06***
41	52	0.8244	0.5527	0.5155
42	9	0.008015**	5.631e-09***	5.596e-10***
42	34	9.426e-07***	0.6607	5.114e-05***
42	56	0.002799 **	0.4676	0.0007514 ***
44	3	0.8185	0.01423*	0.0008259***
44	5	0.000592***	0.4559	0.06925.
44	7	2.13e-05***	0.5735	0.02716*
44	11	0.005773**	0.5936	0.0003953***
44	12	1.425e-05***	0.5647	0.0128*
44	15	5.145e-06***	0.8074	6.366e-08***
44	30	0.01109*	0.4203	0.01448*
45	8	0.0005732***	0.3636	0.003786**
45	18	0.0001402***	0.6531	1.75e-06***
45	27	6.69e-05***	0.4072	3.024e-05***
46	21	4.024e-06***	0.004087**	6.898e-09***
46	28	0.0005179***	0.459	1.291e-05***
47	19	1.971e-09***	0.1338	6.513e-09***
47	20	2.56e-05***	0.8035	1.633e-07***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The adult data consists of all sound tokens that were acquired from every adult speaker in the control group, and those tokens were compared to the sound tokens of each child speaker's production. As shown in Table 3.10, VOT values in the two generation groups for identical stop categories were compared. Significant differences in VOT values in this model indicate the fact that the children's VOT range is significantly different from the adult speakers' VOT range. That is, these were immature productions of stop categories in terms of VOT. Similarly to the previous t-tests, this analysis shows the degree of similarity in VOT variation in the two groups. Even though a clear relationship of age with VOT variation might not be found at a glance, it is interesting to note the pattern that the child speakers showed relatively better achievement with respect to VOT in the three stop categories with a few exceptions; *childid_24* and *childid_31*, aged 27 and 39 months respectively, showed significantly different VOT values in their productions of all three categories compared to those in adult productions. It has been pointed out that the child speakers tend to make a VOT distinction especially for fortis stops in their speech, which was also shown in Table 3.9. However, their actual VOT values are still different from the VOT values in the fortis stops produced by adult speakers. This tendency does not seem closely related to a specific age according to Table 3.10, since this pattern is observed throughout the analysis.

In the case of comparison of lenis or aspirated stops, children aged 42 months or older tend to have VOT values that are statistically similar to those in the adult productions. Relatively little similarity in VOT can be found in the comparison in fortis stops even among the older children. This is somewhat surprising since we have seen the consistent production pattern that the child speakers articulate fortis stops with distinctively shorter VOT values, resulting in more mature production of fortis stops than the other two stop categories. Therefore, the phonetic resemblance in VOT between children and adult speakers can be accounted for through this statistical method, but we should note that a low level of resemblance does not necessarily mean that the distinction in stop contrast is poorly done, as a relative phonetic difference is much more important. Rather, it indicates that those sound tokens would sound immature or not adult-like.

Table 3.10. Results of two-sample t-tests of VOT comparing adult and child data.

Age (in months)	childid	Adults' lenis- child's lenis	Adults' fortis-child's fortis	Adults' asp-child's asp
20	48	0.5269	0.1815	0.06514.
22	49	0.2763	0.06624.	0.3584
24	46	0.01094*	0.08791.	0.07615.
24	58	< 2.2e-16 ***	0.9463	1.312e-06 ***
26	26	0.8133	0.02293*	0.2326
26	40	0.00436**	0.1959	0.2788
27	24	0.003656**	0.003167**	0.0001157***
27	39	0.8506	0.01821*	0.8223
27	42	0.7029	0.08401.	0.1148
29	44	0.1413	0.05367.	0.01272*
29	45	0.8318	0.002476**	0.3531
30	23	0.2807	0.06214.	0.01767*
30	38	0.3783	0.01913*	0.8717
30	41	0.5474	0.01311*	0.04646*
31	57	0.3425	0.03507 *	0.06561 .
32	25	0.3627	0.3573	0.2101
32	43	0.04431*	0.03678*	0.6557
32	47	0.1497	0.06412.	0.2765
32	53	0.5136	0.7155	0.7035
33	22	0.7983	0.1685	0.05161.
33	55	0.5307	0.004353 **	0.502
34	50	0.8931	0.01041*	0.09418.
34	51	0.3136	0.00787 **	0.774
34	54	0.002678 **	0.225	0.2198
35	4	0.07607.	0.2076	0.03949*
35	13	0.6663	0.04128*	0.9021
36	16	0.9767	0.1955	0.7177
37	33	0.7013	0.01399*	0.6292
38	1	0.0008163***	0.1465	0.007482**
38	17	0.001899**	0.02752*	0.6657
38	35	0.1503	0.1394	0.1699
38	36	0.01681*	0.01954*	0.9858
39	2	0.00382**	0.2701	0.007279**
39	6	0.02411*	0.0005765***	0.4726
39	10	0.002129**	0.116	0.1391
39	31	0.01437*	0.002196**	0.001427**
39	32	0.07267.	0.02564*	0.9797
39	37	0.05654.	0.2192	0.1287
40	14	0.893	0.008785**	0.9264
40	29	0.1732	0.004015**	0.0979.
41	52	0.01354 *	0.01497 *	0.526
42	9	0.8224	0.2554	9.22e-08***
42	34	0.8812	0.06903.	0.007452**
42	56	0.2255	0.01436 *	0.5551
44	3	0.3235	0.1413	0.7929
44	5	0.526	0.02077	0.6472
44	7	0.4173	0.5275	0.1958
44	11	0.1241	0.007484**	0.9225
44	12	0.01945*	0.324	0.9833
44	15	0.07946.	0.06248.	0.7461
44	30	0.5691	0.02771*	0.8899
45	8	0.09349.	0.0929.	0.135
45	18	0.2096	0.2614	0.8416
45	27	0.9537	0.249	0.0007189***
46	21	0.5691	0.02771*	0.8899
46	28	0.2134	0.04207*	0.6339
47	19	0.0008307***	0.9324	0.2063
47	20	0.04195*	0.19	0.9497

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

From the results of the two-way t-tests, phonetic resemblance in VOT can be observed. The statistical analysis for the adult data showed that adult speakers tend to have significant VOT differences even between lenis and aspirated stops, but the correlation between fortis and VOT is clearer even in children's productions. This finding also implies that even young children are able to articulate fortis stops with appropriate VOT distinction. This also supports Kong's study (2009), which argues that fortis stops are mastered earlier than the other two stop categories because of their VOT range.[3] Given that VOT values in fortis stops are distinctively different from those in lenis or aspirated stops in children's production, it is notable that as VOT is a universally used phonetic feature, it is easier for young children to acquire or master. The results of the t-tests comparing the child and adult data can give us an impression that all child participants generally did a good job in differentiating the three stop categories in their speech in spite of a few outlying cases. However, a clear relationship between age and mature linguistic performance in terms of VOT regarding stop contrasts is hardly found. The t-test comparison with the adult data does not show a certain clear-cut age by which VOT is completely acquired or mastered.

This unclear result itself is linked to the fact that language acquisition and linguistic competence are affected by children's inborn abilities and linguistic surroundings, so a specific age cannot guarantee a certain linguistic performance or capacity, and children's linguistic performance can differ depending on various factors.

---

[3] Kong (2009) used the criterion of mastery as Amayreh & Dyson (1998), Smit *et al.* (1990), and So & Dodd (1995) suggested. According to them, the definition of 'mastering' of a specific phonation type is determined as over 75% of the production tokens meeting the given phonetic standard. The definition of 'acquisition' of a specific phonation type means that over 50% of the production tokens meet the phonetic standard.

In spite of this fact, it is still interesting that even younger children, who are under 3 years of age, were aware of the role of VOT and actively produced it especially for fortis stops. This means that at an early age, toddlers have acquired and developed their native acoustic parameter in their production system.

Another acoustic parameter, F0, is used as a language-specific acoustic cue to distinguish Korean stop categories (Jun, 1993; Jun, 1996). Due to this tendency, it has been shown that F0 is acquired later than VOT, and that its role is minimal even in perceptual Korean stop distinction (e.g., Kong, 2009; Kong *et al.*, 2011). In order to examine the role of F0 in distinction between lenis and aspirated stops, and to understand the developmental patterns related to it, the two-sample t-test was performed again.

In order to determine whether the child speakers showed a significant phonetic difference between the two different stop groups in terms of F0, the t-test was conducted. Each child speaker's F0 values were compared in three pairs, lenis-fortis stops, lenis-aspirated stops, and fortis-aspirated stops. Generally, the results of the t-tests show that the child participants did not have significant F0 differences between the stop contrasts, and the phonetic differences are relatively much smaller compared to the results of the VOT comparison. In particular, in the case of the pair of fortis and aspirated stops, the children were not able to produce a significant F0 difference in general. Only seven children showed significant F0 differences between fortis and aspirated stops (when p-values < 0.05,), and their ages were all different as they are ranged from 27 months to 47 months.

In the pair of lenis-fortis stops, even the relatively older children aged 44 months to 47 months were not able to differentiate those stop categories in F0 at the following vowel onset. The observed inconsistency in the F0 comparison leads us to think that the relationship between the children's ages and their linguistic achievement are not that close. For example, some 38-month- to 42-month-old children tended to produce a significant F0 difference between lenis and fortis stops, while 45 or 46 month-old children failed to show any significance in their F0 differences.

In the case of lenis-aspirated stops, however, it seems that children's age is positively related to the acquisition of F0. Relatively older children showed that they produced significant F0 differences between these two stop categories compared to the much younger children. For example, before 31 months, no significant F0 difference in this pair of stop categories was found except for one child (*childid_48*, the youngest child, aged 20 months). After 44 months, the children's F0 distinctions were generally significant (p-values < 0.05).

Table 3.11. Result of two-sample t-tests for F0 in child productions of three different stop categories.

Age (in months)	childid	Lenis-Fortis	Lenis-Aspirated	Fortis-Aspirated
20	48	0.2936	0.01381*	0.0624.
22	49	0.1733	0.6588	0.1284
24	46	0.7056	0.5144	0.7956
24	58	0.02937 .	0.1307	0.8998
26	26	0.1343	0.1	0.6408
26	40	0.1276	0.2557	0.6589
27	24	0.0333*	0.1384	0.6814
27	39	0.0516.	0.9436	0.04641*
27	42	0.7865	0.8886	0.6966
29	44	0.9848	0.7305	0.7731
29	45	0.05021.	0.5857	0.09315.
30	23	3.158e-06***	0.1769	0.03571*
30	38	0.9191	0.5221	0.7061
30	41	0.3072	0.07564.	0.234
31	57	0.3341	0.5241	0.9868
32	25	0.04247*	0.003526**	0.5597
32	43	0.9271	5.608e-05***	0.001772**
32	47	0.06457.	0.5494	0.2637
32	53	0.009916**	0.001047**	0.4052
33	22	1.991e-05***	0.0004291***	0.5827
33	55	0.2008	0.505	0.4769
34	50	0.8163	0.374	0.5634
34	51	0.4368	0.01176*	0.05158 .
34	54	0.07234 .	0.9685	0.122
35	4	0.06988.	0.1844	0.7764
35	13	0.06638.	0.01777*	0.8802
36	16	0.1003	0.2364	0.5784
37	33	0.004442**	0.0007759***	0.2053
38	1	0.497	0.06141.	0.2529
38	17	0.7932	0.5179	0.1432
38	35	0.03175*	0.2972	0.4368
38	36	0.003796**	0.6245	0.07486.
39	2	0.00557**	0.00541**	0.6536
39	6	0.2634	0.7951	0.4947
39	10	0.0006497***	0.03903*	0.09298.
39	31	0.07729.	0.5036	0.4856
39	32	0.1239	0.3248	0.7392
39	37	0.001698**	0.01728*	0.2898
40	14	0.00273**	1.49e-05***	0.04133*
40	29	0.03334*	0.1238	0.7044
41	52	0.8962	0.8822	0.5685
42	9	5.765e-06***	1.041e-10***	0.961
42	34	0.305	0.3147	0.08949
42	56	0.4523	0.8956	0.5028
44	3	0.05436.	0.3071	0.4414
44	5	0.01342*	0.2993	0.236
44	7	0.3232	0.3017	0.933
44	11	0.7266	0.1154	0.3906
44	12	0.01544*	0.003591**	0.9666
44	15	0.003041**	0.0003592***	0.7844
44	30	0.7878	0.4596	0.5975
45	8	0.5187	0.7589	0.9
45	18	0.3363	0.008742**	0.05603.
45	27	0.2583	0.007048**	0.02623*
46	21	0.7878	0.4596	0.5975
46	28	0.6276	5.096e-06***	0.02079*
47	19	0.03298*	3.737e-06***	0.001028**
47	20	0.06366.	0.00102**	0.05544.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The child speakers' articulatory distinctions do not resemble the control groups' F0 differentiation in that the adult speakers showed significant F0 differences in every stop contrast. For example, a 44-month- and 47-month-old child (*childid_18* and *childid_20*, respectively) phonetically differentiated lenis and aspirated stops, whereas they did not produce a significant F0 difference in the other stop pairs. Only one child (*childid_19*) showed a significant F0 difference in every stop contrast. Most of the children aged 20 to 31 months failed to produce a significant F0 difference in any stop contrast. The observed difficulty to produce F0 differences allows us to consider the order of acquisition of VOT and F0, in which children acquire VOT prior to F0.

The weakest F0 distinction can be found in the comparison of fortis-aspirated stops. We assume that VOT can solely play a crucial role to define fortis stops, so F0 distinction between fortis and aspirated stops seems less motivated than between lenis and aspirated stops. Note that the F0 distinction is significantly different in all pairs in the adult productions, so even the oldest children here would not have mature stop distinction in F0. This observation would not seem to concur with the claim that all of the different stop categories are acquired or mastered before 3 years of age, which has been reported in previous transcription-based studies (e.g., Y. T. Kim, 1996; M. J. Kim & Pae, 2005; Um, 1994). Since consistent phonetic differentiation is not shown even after 36 months, we still need a more detailed statistical model to test the effect of age on the acquisition of F0 and on the development of phonetic categorization of stop phonation types.

3.6.2.2.2. Simple regression model

As the previous t-tests were limited in not being able to show a clear correlation between children's age and the development of VOT or F0, a simple regression model was designed. In this model, VOT values or F0 values should be dependent variables which are affected by the main factor, age. However, a linear correlation between VOT or F0 values and age cannot be simply interpreted; absolutely large or small values of such phonetic measures do not mean successful mature productions. The maturity of the productions in terms of VOT or F0 can be judged by the similarity with the control group's VOT or F0 values. Therefore, the absolute VOT or F0 difference between the child and adult speakers' productions is the optimally appropriate dependent variable in this regression model and it suggests a simple equation; the bigger the absolute differences, the less mature the productions. In this analysis, the three different stop categories were dealt with separately since VOT and F0 differences differ depending on which stop groups are compared. Therefore, an identical model was conducted three times. It is notable that average values in productions of each POA in each stop category by each participant were compared to the control group's equivalent (the identical POA in the same stop category) average VOT/F0 values. The use of nine averaged (labial, coronal, velar) × (lenis, fortis, aspirated stops) values by each child participant is to prevent putting too much weight on any specific age with a relatively large number of sound tokens in the analysis. After calculating the differences from the adult speakers' average VOT values, all of the obtained differences were converted to absolute values.

These absolute differences in VOT are the dependent variable while age and gender are used as independent variables in the model.

In this process, raw VOT values in each stop category could be used. However, F0 differences as a dependent variable in this model needed to undergo some manipulation. Since F0 is a relative acoustic measure, the mean F0 values of the control group were not used as they are. Instead, I decided to use the relative differences between different stop categories such as F0 difference between aspirated and lenis stops, in order to standardize the F0 values by different age or gender groups. Such F0 differences between different stop groups were calculated only within the productions by the same gender (male/female) or age (adult/child) groups. Although F0 is a relative acoustic measure, the F0 differences can be standardized this way, as F0 differences between stop categories should be alike in spite of the different speaker groups. As previously shown in the adult productions, F0 values for lenis stops tend to be the lowest, while those for aspirated stops are likely to be the highest, so the F0 differences from aspirated to fortis or lenis stops were preferred in order to avoid getting negative F0 difference values. The three different pairs with three stop categories such as aspirated-lenis stops, aspirated-fortis stops, and fortis-lenis stops, were used to calculate F0 differences in the same gender adult group.

For this simple regression model, male adult speakers' F0 values were provided and the F0 differences across the three stop categories were calculated. After getting the standardized F0 differences in three different pairs of stops, F0 differences by each child in the identical pair were deducted from the standard F0 values. As previously mentioned, in order to prevent the unintended effects of the unbalanced number of tokens on the results, the mean F0 values in each speaker's productions were used so the same numbers of F0 values could be analyzed. Through this process, the acquired F0 differences became a dependent variable and child speakers' age is the independent variable.

The first simple regression model tested the relationship between VOT differences in lenis stops and age. The results show that age and the adult-like VOT values in lenis stops are not significantly related ($p > 0.05$, *coef.* $= -0.023$). The regression curve on the scatterplot in Figure 3.20 represents how the correlation curve can be drawn between age and the VOT differences from adult speakers' productions. The negative coefficient for *Age* makes the line slope downwards slightly. However, this relationship seems quite negligible since it is not statistically significant.

Table 3.12. The output of the simple regression model for VOT differences and age in lenis stops.

Coefficients:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	22.76641	7.89023	2.885	0.00441 **
Age	-0.02307	0.21362	-0.108	0.91411

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

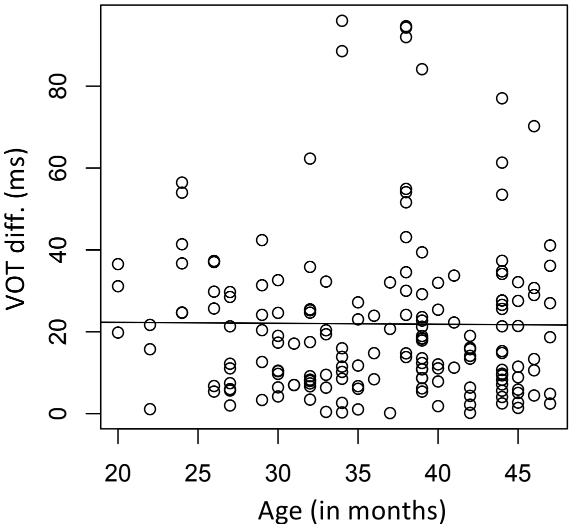

Figure 3.20. The simple regression line on a scatterplot for VOT differences between children's and adults' productions of lenis stops.

An identical regression design was applied to VOT differences in fortis stops. There is a tendency for relatively older children to produce fortis stops in a more adult-like manner in terms of VOT (*coef.* = −0.1087), but this tendency is not significantly related to age (*p* > 0.05). The figure below also represents this relationship with a regression line, which is statistically insignificant.

Table 3.13. The output of the simple regression model for VOT differences in fortis stops.

Coefficients:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	14.3352	3.8918	3.683	0.000308 ***
Age	-0.1087	0.1054	-1.032	0.303607

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

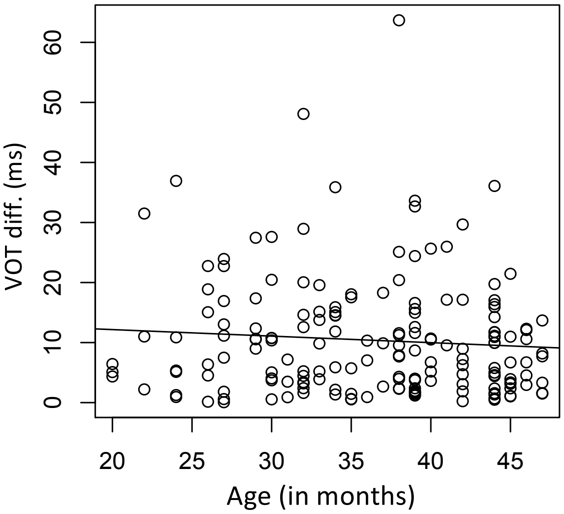

Figure 3.21. The simple regression line on the scatterplot for VOT differences between children's and adults' productions of fortis stops.

The output of the simple regression model for aspirated stops illustrates that *Age* is significantly related to VOT difference (*coef.* = −0.7724, $p < 0.05$). As the figure shows, VOT values of aspirated stops produced by the older children are more adult-like and mature than those produced by the younger children. The coefficient value is the lowest among the three stop categories, so the regression line slopes downward more sharply. This slope implies that the VOT differences decrease as the speakers' age increases.

Table 3.14. The output of the simple regression model for VOT differences in aspirated stops and age.

Coefficients:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	53.2688	9.3041	5.725	4.51e-08 ***
Age	-0.7724	0.2519	-3.066	0.00252 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

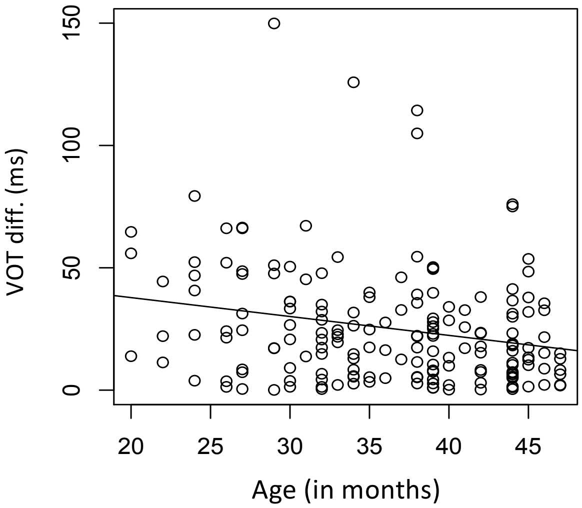

Figure 3.22. The simple regression line on the scatterplot for VOT differences between children's and adults' productions of aspirated stops.

Through the simple regression model, I have found that there is a negative correlation between children's age and VOT differences from the adult speakers' productions. That is, older children tend to have more adult-like and mature VOT production than younger children, since the coefficients of *Age* are all negative values. A surprising finding is that the significance level for the acquired coefficients is the highest for aspirated stops. Generally, it has been shown that fortis stops are more easily acquired at an early age than the other stop manners. This fact allows us to consider the possibility that a close relationship between age and VOT differences in fortis stops could not be found simply because even the youngest children are already producing quite mature VOT values in that stop category. That is, the developing pattern with a decreasing line would hardly be found if all the children performed as well as the adult speakers.

Similar to the case of fortis stops, lenis stops have not shown any significant relationship of age to adult-like VOT productions. This observation would imply two possibilities. First, there could be a non-crucial role of VOT in production of lenis stops even in the case of the control group. Generally the phonetic definition of lenis stops involves a certain combination of VOT and F0 ranges, but here the role of F0 would be much more determinant (Kim, 2008; Kong, 2009). As we recall the results of the control group, the actual VOT values in the lenis productions were widely dispersed compared to the other stop categories. Considering this phonetic characteristic of lenis stops, the weakest relationship between age and VOT differences is quite understandable. Another speculation is related to the difficulties in the acquisition of lenis stops. Acoustic implementation of lenis stops seems to be relatively difficult to learn and apply because of their intermediate VOT values. The difficulties in the appropriate acoustic implementation would make lenis stops perceptually less salient (e.g., Kim, 1965). Therefore, even the older children could be unable to articulate lenis stops, so any significant developmental pattern would not be found. A significant effect of age has been found in the production of aspirated stops, so it still suggests that VOT is a crucial acoustic parameter to develop between the age of 2 and 4 years in the acquisition of stop contrasts in Korean.

On the other hand, a developmental pattern in F0 was found through an identical simple regression model, in which the F0 difference between adult and child speakers' productions is the dependent variable. Using the three different kinds of F0 differences (between aspirated and lenis stops, aspirated and fortis stops, and fortis and lenis stops), this model investigates if there is a significant relationship between age and F0 development in production of stops. The first model was performed to see the F0 differences between aspirated and lenis stops. According to the output, the effect of *Age* on F0 difference between aspirated and lenis stops is significant (*coef.* = −1.1889, $p <$ 0.05). The coefficient of *Age* is negative, which indicates that older children tend to have a smaller F0 difference, so they have more adult-like F0 differences between aspirated and lenis stops.

Table 3.15. The output of the simple regression model for F0 differences between aspirated and lenis stops.

Coefficients:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	78.7152	20.2688	3.885	0.000273 ***
Age	-1.1889	0.5485	-2.167	0.034473 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

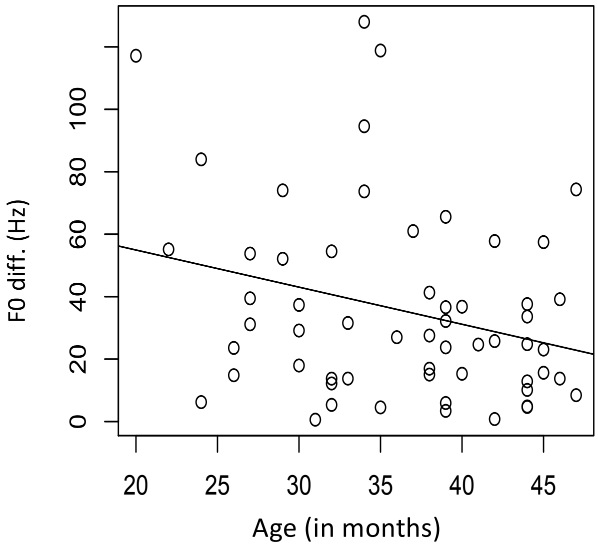

Figure 3.23. The simple regression line on the scatterplot for F0 differences between aspirated and lenis stops according to children's age.

The associated figure illustrates how children's age relates to F0 differences between aspirated and lenis stops. The slope is negative, which means the productions of older children are likely to have a smaller F0 difference from the production of the control group. This implies that between the ages of 2 and 4 years, in the acquisition of lenis and aspirated stops, child production of the acoustic implementation using F0 is significantly changed in a 'mature' way. F0 begins to function as a useful tool in differentiating stop contrasts. As the control group's productions have shown, the F0 differences between aspirated and lenis stops were the biggest, so this regression model indicates that the children have learned to make a substantial F0 difference for producing aspirated or lenis stops.

The second comparison is the pair of aspirated and fortis stops. In this comparison, the effect of *Age* on the F0 difference is negligible because the obtained *p*-value for the coefficient is high enough (*coef.* = −0.7573, *p* > 0.1). This means that the change in F0 differences according to age seems insignificant. The coefficient is a negative value, which indicates that the older children's F0 differences between aspirated and fortis stops have become more adult-like, though this is not a significant tendency.

Table 3.16. The output of the simple regression model for F0 differences between aspirated and fortis stops.

Coefficients:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	64.8827	18.9376	3.426	0.00115 **
Age	-0.7573	0.5127	-1.477	0.14526

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

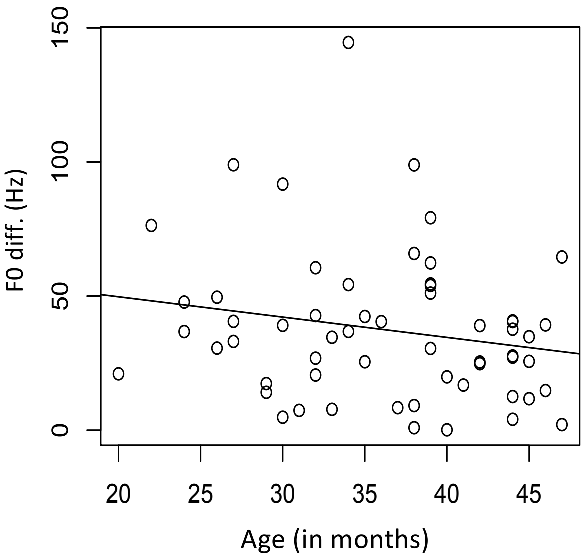

Figure 3.24. The simple regression line on the scatterplot for F0 differences between aspirated and fortis stops according to children's age.

Lastly, an identical regression analysis is done for the F0 difference between fortis and lenis stops. The results show that *Age* is somewhat negatively related to F0 differences between fortis and lenis stops, but the correlation is not strongly significant (*coef.* = −0.7792, $0.05 < p < 0.1$).

Table 3.17. The output of the simple regression model for F0 differences between fortis and lenis stops.

Coefficients:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	62.7971	16.8923	3.717	0.000467 ***
Age	-0.7792	0.4573	-1.704	0.093945 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

This associated figure illustrates how *Age* is related to the F0 difference between fortis and aspirated stops. The reported coefficient is negative, which implies that the phonetic difference from the adult productions would be likely to decrease with age. The significance level is not high enough to pay much attention to this tendency, but it possibly suggests that age effects could be found in the process of acquisition of F0 as a useful tool for the fortis-lenis stop distinction.

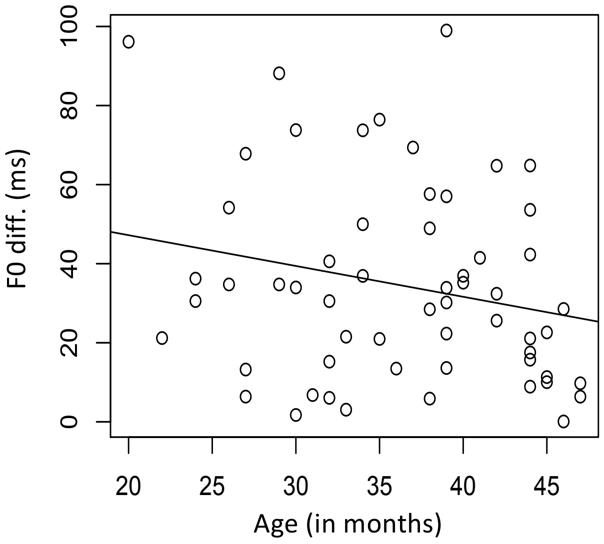

Figure 3.25. The simple regression line on the scatterplot for F0 differences between fortis and lenis stops according to children's age.

In the three different simple regression models, all the reported coefficients are negative values. This implies that younger children tend to have a hard time producing F0 differences for articulating different stop categories, while the older children's production patterns resemble the adult speakers' distinctions between two stop categories. However, if we consider the degree of significance in each case, the only significant development was found in the case of the phonetic difference between aspirated and lenis stops. As we recall, in the results of the t-tests by the control group, the F0 values in the three stop categories were significantly different (*p*-values < 0.05). Compared to this, the child speakers' F0 differences are not big enough to differentiate the stop contrasts. In particular, F0 differences of fortis-lenis stops or aspirated-fortis stops were not likely to become mature enough to be judged to be successful distinction. This might be brought out by the possibility that fortis and lenis stops have a relatively subtle phonetic difference so that language-learning toddlers cannot reach certain phonetic standards. As mentioned previously, the biggest benefit to using F0 as an acoustic tool is to distinguish lenis and aspirated stops, so the relatively poor phonetic distinction between other stop categories seems quite understandable.

To summarize, the only thing to clearly note here is that older children tend to differentiate lenis and aspirated stops relatively successfully with big enough F0 differences. The degrees of F0 difference are significantly affected by children's ages, so depending on age, a clear development in using F0 to distinguish lenis and aspirated stops was observed. The lack of evidence of significant change in F0 differences according to age in the other two stop pairs allows us to consider the possibility that development of F0 as an acoustic parameter of Korean stop contrast has not been completed between the ages of 2 and 4 years.

3.6.2.2.3. A mixed-effects linear regression model

The analysis through the simple regression model used an absolute difference in VOT and F0. It also allowed us to draw a regression line, which represents a relationship between children's ages and language competence in the acoustic implementation of

Korean stop contrasts in terms of VOT and F0. However, this simple regression analysis seems to simplify the complexity of the child data mainly because it analyzes only average values. Using an unbalanced number of tokens in the simple regression model puts unnecessary emphasis on the older children's results, which have relatively larger numbers of sound tokens. In a mixed-effects model, the unbalanced number of tokens child by child can be handled, as the model automatically estimates the error terms, which represent an unobservable individual speaker-difference and a production-difference. As mentioned earlier, it is extremely important to use all the production tokens in a child language study, as each trial conveys an important implication of how these children's language development occurs given that their productions vary depending on the given contexts. A mixed-effects linear model is considered the most useful and appropriate method for overcoming these deficits. Through a mixed-effects linear regression model, the acoustic properties of the stop categories and the developmental patterns in the use of VOT and F0 can be observed. Applying this multi-level model to the child data, there is no need to compare to the phonetic standards supplied by the control group. The two representative acoustic parameters, VOT and F0, will be dealt with separately in relation to other predictor variables. For the analysis, the *lme4* package (Bates *et al.*, 2011) was implemented in R (R Development Core Team, 2011) and the *lmer* function was used. In this multi-level regression, VOT (ms) and F0 (Hz) are the dependent variables and by-speaker adjustment to intercept is a random effect.

First, the results from the mixed-effects regression model for VOT are represented. Before applying more complex interactions, in order to figure out the independent effects of each variable, the simplest linear regression model is designed as in (4) below.

(4) $VOT_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Male_i + \beta_3 \times Fortis_{ij} + \beta_4 \times Asp_{ij} + \eta_i + \varepsilon_{ij}$

(Reference categories: *Female* for *Gender*, *Lenis* for *Stop category*)

This model was run to see how each fixed effects predictor independently affects VOT.[4] Gender should have binary values, male or female, so the variable *Female* was used as a reference term. Stop category has three different variables, so only two terms were used in the equation; *Lenis* was the reference category. The index *i* represents an individual level. Since age, gender, and the error term η vary depending on who was the speaker, the three variables have the same index *i*. Another index *j* represents a production level, so other variables that are directly related to a production are *ij*-indexed. The most crucial benefit of the hierarchical model is that it includes error terms η and ε, which are on different levels. The core of the equation is the error term η, which explains unobservable individual differences. That is, it can represent so-called "inborn talent" such as superior linguistic ability. Linguistic ability should differ child by child, so with this equation, the hierarchical model is able to take into account the differences in linguistic abilities that exist among the child speakers. The second term ε represents a production level error

---

[4] Whether POA could be an independent variable in this hierarchical model should be carefully considered in order to generate reliable results. Generally, labial stops do not guarantee shorter VOT values rather than velar stops across the stop categories, so I decided not to consider how POA independently affects VOT, and did not use it as an independent variable in the linear mixed model here, even though it can be one of the useful interaction terms later.

term, which includes unobservable phonetic differences among productions even by the same speaker. The reference categories and these error terms will be used identically in following equations in this section.

As shown in Table 3.18, the general pattern of VOT variation depending on the three stop categories in the child data can be recognized. The regression model based on (4) shows that fortis stops have significantly shorter VOT values than lenis or aspirated stops (*coef.* = −47.9565, $p < 0.001$). VOT values in aspirated stops are also slightly but significantly longer than those in lenis stops (*coef.* = 7.8771, $p < 0.001$). The effects of age and gender on the VOT ranges are not clear here. This unclear correlation seems quite consistent with the intuitive impression about the relationship between gender or age and VOT. The acquired sound tokens in three different stop categories were used as a dependent variable in the data matrix, so it is fully understandable that older speakers cannot guarantee longer VOT values or that a specific gender is not related to longer VOT values. Therefore, it seems natural that no significant relationship between age or gender and longer VOT values was found. Through the equation in (5), I was able to observe the general VOT patterns over age depending on stop category.

Table 3.18. The output of the mixed-effects linear model based on (4).

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	136.2	11.67	
Residual		1027.9	32.06	
Number of obs: 2331, groups: childid, 58				
Fixed effects:				
	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	57.2607	9.0895	6.3	4.56e-08 ***
Fortis	-47.9565	1.8335	-26.155	< 2e-16 ***
Asp	7.8771	1.8273	4.311	1.71e-05 ***
Age	0.3835	0.2345	1.635	0.108
Male	2.5755	3.5265	0.73	0.468

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The next regression model from (5) was conducted in order to see the correlation between age and stop category in VOT.

(5) $VOT_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Fortis_{ij} + \beta_3 \times Asp_{ij} + \beta_4 \times Fortis_{ij} \times Age_i + \beta_5 \times Asp_{ij} \times Age_i + \eta_i + \varepsilon_{ij}$

(Reference category: *Lenis* for *Stop category*)

Table 3.19. The output of the mixed-effects linear model based on (5).

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	136.6	11.69	
Residual		1022.8	31.98	
Number of obs: 2331, groups: childid, 58				
Fixed effects:				
	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	46.49544	10.60899	4.383	2.77e-05 ***
Fortis	-19.21373	9.63738	-1.994	0.04633 *
Asp	11.37443	9.47566	1.200	0.23014
Age	0.67075	0.27678	2.423	0.01706 *
Fortis:Age	-0.77573	0.25532	-3.038	0.00241 **
Asp:Age	-0.09495	0.25032	-0.379	0.70451

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The fixed effects coefficient of the interaction between *Fortis* stops and *Age* is the most significant ($p < 0.01$) while the interaction between *Aspirated* stops and *Age* is the weakest ($p > 0.1$). With the fixed effects coefficients and intercept, the following figure illustrates how VOT changes in the production of three different stop categories depending on child speaker's age.[5]

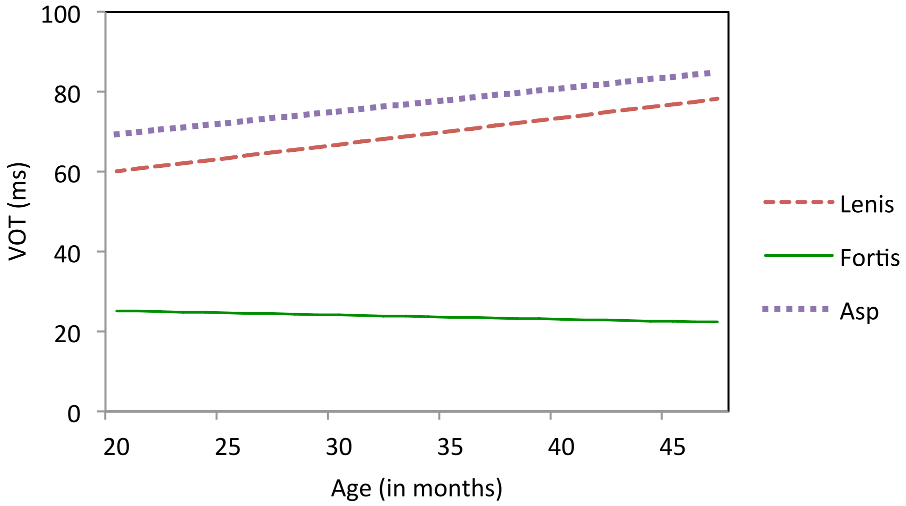

Figure 3.26. Interaction effects of stop categories and age (in months) on VOT (ms).

The correlation lines in the figure above indicate that in older children's articulations VOT values in lenis stops tend to be longer while VOT values in fortis stops tend to be shorter. This pattern implies that the VOT difference between lenis and fortis stops has increased enough to distinguish those two phonation types as the children are growing and acquiring language. The interaction between VOT and lenis or fortis stops is found to

---

[5] According to Table 3.19, the expected VOT changes depending on speaker's age (in months) should be yielded as follows:

VOT change in lenis stop = 46.49544 + (0.67075×Age)
VOT change in fortis stop = −19. 21373+46.49544 + ((−0.77574+0.67075)×Age)
VOT change in aspirated stop = 11.37443 + 46.49544 + ((−0.09495+0.67075)×Age)

be significant (p-values < 0.05) in the target-aged children, who are 2 to 4 years old. This implies that during the period of 2 to 4 years, a significant VOT development in relation to Korean stop contrast takes place.

To see if there is any gender effect, another mixed-effects model with a different interaction term was conducted.

$$(6)\ VOT_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Male_i + \beta_3 \times Fortis_{ij} + \beta_4 \times Asp_{ij} + \beta_5 \times Fortis_{ij} \times Male_i + \beta_6 \times Asp_{ij} \times Male_i + \eta_i + \varepsilon_{ij}$$

(Reference categories: *Lenis* for *Stop category, Female* for *Gender*)

Table 3.20. The output of the mixed-effects linear model based on (6).

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	135.3	11.67	
Residual		1028.6	32.07	
Number of obs: 2331, groups: childid, 58				
Fixed effects:				
	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	57.8224	9.1451	6.323	3.84e-08 ***
Fortis	-48.3284	2.3272	-20.767	< 2e-16 ***
Asp	6.6271	2.337	2.836	0.00462 **
Age	0.3833	0.2346	1.634	0.10806
Male	1.178	4.1613	0.283	0.77767
Fortis:Male	0.9189	3.7824	0.243	0.80807
Asp:Male	3.2232	3.7508	0.859	0.39027

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

As the fixed effects predictors reveal, any significant interaction between gender and stop category was not found, which means gender difference does not affect VOT change in production depending on three different stop categories. This is quite compatible with what was expected, but it needs to be determined if female children’s performances could be superior to male children's productions, since female children tend to learn their native language faster than male children. This model confirms that there is no significant gender difference in VOT distinction for stop contrasts.

Next, in order to determine whether POA can significantly affect VOT changes within the same stop category, the multi-level regression model in (7) includes related interaction terms such as *STOP category* × *POA*.

(7) $VOT_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Male_i + \beta_3 \times Fortis_{ij} + \beta_4 \times Asp_{ij} + \beta_5 \times Alveolar_{ij}$
$+ \beta_6 \times Velar_{ij} + \beta_7 \times Fortis_{ij} \times Alveolar_{ij} + \beta_8 \times Fortis_{ij} \times Velar_{ij}$
$+ \beta_9 \times Asp_{ij} \times Alveolar_{ij} + \beta_{10} \times Asp_{ij} \times Velar_{ij} + \eta_i + \varepsilon_{ij}$

(Reference categories: *Lenis* for *Stop category, Female* for *Gender*, *Labial* for *POA*)

Table 3.21. The output of the mixed-effects linear model based on (7).

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	137.5	11.72	
Residual		1012	31.81	
Number of obs: 2331, groups: childid, 58				

Fixed effects:				
	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	59.6628	9.2393	6.458	2.04e-08 ***
Fortis	-53.0634	2.7984	-18.962	< 2e-16 ***
Asp	10.4221	2.9032	3.59	0.00034 ***
Age	0.3949	0.2351	1.68	0.09858 .
Male	2.4205	3.5349	0.685	0.49642
Alveloar	-8.8663	3.3102	-2.678	0.00746 **
Velar	-1.5213	3.0136	-0.505	0.61374
Fortis:Alveolar	5.9923	4.9433	1.212	0.22559
Asp:Alveolar	-3.9899	4.5433	-0.878	0.37995
Fortis:Velar	8.6628	4.0989	2.113	0.03470 *
Asp:Velar	-3.3038	4.2525	-0.777	0.4373

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The results of the model in (7) show that generally POA and VOT values in any stop category are not closely related. However, a significant VOT change is found between alveolar lenis stops and labial lenis stops (*coef.* = -8.8663, $p < 0.01$). A significant correlation is also found when fortis stops are produced with velar POA (*coef.* = 8.6628, $p < 0.05$). It is important to note that in production of fortis stops, there is a significant relation to POA, which can be interpreted as indicating that the acquisition of fortis stops occurs prior to the other stop types, so the acoustic implementation of VOT would be further specified, depending on POA.

On the other hand, it is interesting to note the idiosyncratic production pattern that in the case of lenis stops, VOT values in labial stops are longer than alveolar or velar stops. This pattern is not quite consistent with the well-known adult production pattern in which velar stops have longer VOT values than alveolar or labial stops within the identical stop category. Therefore, this pattern itself indicates the immature production of lenis stops by the child speakers. Even though their articulatory distinction between three different stop groups seems quite robust at a glance, the significant effects of POA on stop production cannot be found in general.

The following analysis deals with the correlation between the predictors and another acoustic parameter, F0 as the response variable. As we recall, the simple regression model revealed that there is some significant relationship between age and the development of F0. In particular, successful phonetic differentiation between lenis and aspirated stops is closely related to children's age in that model. This means that F0 narrowly functions for phonetic discrimination between aspirated and lenis stops by young children.

The multi-level regression model for F0 was designed based on the same structure used in the VOT analysis. The simplest model includes all factors as fixed effects predictors, with F0 as the only dependent variable as follows.

(4′) $F0_{ij} = \beta_0 + \beta_1\times Age_i + \beta_2\times Male_i + \beta_3\times Fortis_{ij} + \beta_4\times Aspirated_{ij} + \eta_i + \varepsilon_{ij}$

(Reference categories: *Lenis* for *Stop category, Female* for *Gender*)

Table 3.22. The output of the mixed-effects linear model based on (4′).

Random effects:

Groups	Name	Variance	Std. Dev.
childid	(Intercept)	947.2	30.78
Residual		5114.6	71.52

Number of obs: 2331, groups: childid, 58

Fixed effects:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	351.0342	23.2147	15.121	< 2e-16 ***
Fortis	34.4248	4.0926	8.412	< 2e-16 ***
Asp	38.7306	4.0783	9.497	< 2e-16 ***
Age	-1.1092	0.6004	-1.847	0.0699 .
Male	-25.2119	9.0357	-2.79	0.0072 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

According to the results, the three stop categories are significantly related to F0 in the children’s speech. There is no interaction term included, so the fixed effects coefficients represent the effect of each predictor on F0 values when other variables equal zero. The effect of age on F0 values is not significant, which is naturally understandable since the correlation between child speakers’ age and overall higher or lower F0 values has no important implication. Therefore, I tried to include one interaction term for stop category and age using the equation (5′). It was built to confirm how closely age is correlated with acoustic distinction in F0 between two different stop categories. The interaction term *Stop category*×*Age* was included in the following equation.

$$(5') \; F0_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Fortis_{ij} + \beta_3 \times Aspirated_{ij} + \beta_4 \times Fortis_{ij} \times Age_i + \beta_5 \times Aspirated_{ij} \times Age_i + \eta_i + \varepsilon_{ij}$$

(Reference categories: *Lenis* for Stop category*, Female* for Gender)

Table 3.23. The output of the mixed-effects linear model based on (5′).

Random effects:

Groups	Name	Variance	Std. Dev.
childid	(Intercept)	1067	32.67
Residual		5081	71.31

Number of obs: 2331, groups: childid, 58

Fixed effects:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	364.4817	26.2773	14.43	< 2e-16 ***
Fortis	31.06301	22.5122	1.367	0.1719
Asp	-30.2326	22.1354	-1.385	0.1661
Age	-1.75223	0.74297	-2.358	0.0205 *
Fortis:Age	0.09145	0.60721	0.151	0.8803
Asp:Age	1.88472	0.5929	3.179	0.0015 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Interestingly, the results reveal that the effects of *Age* are found in the relationship with lenis stops ($p < 0.05$) and aspirated stops ($p < 0.01$), but not with fortis stops ($p > 0.1$). According to the fixed effects coefficients, lenis stops tend to have lower F0 values (*coef.* = −1.75223) while aspirated stops tend to have higher F0 values by 0.13249 Hz (= −1.75223 + 1.88472) as age increases by 1 month (p-values < 0.05). This production pattern means that the phonetic gap, in terms of F0, becomes larger with age. This tendency would imply that the acoustic differentiation between lenis and aspirated stops becomes more specific and more accurate as children get older. The following figure illustrates the results of the mixed effects model (5′) with regression lines.[6]

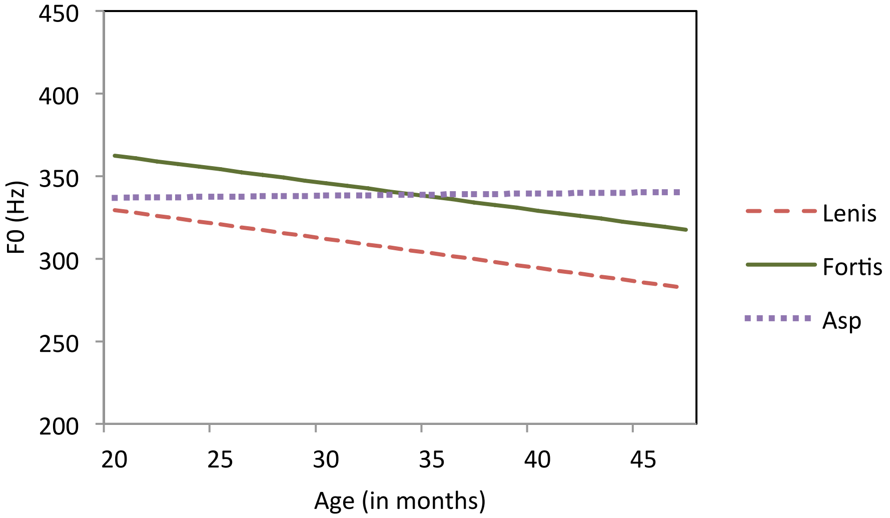

Figure 3.27. Interaction effects of stop categories and age on F0 (Hz).

As shown in the figure above, the F0 change over age has a significant phonetic difference, especially between lenis and aspirated stops. Around 20 months, the F0 values in production of lenis and aspirated stops are almost identical, which means toddlers of 20 months would not differentiate those stop categories using F0. However, this tendency does not persist as they get older. The children begin to show substantial F0 differences between the two stop groups. Around 35 months, fortis and aspirated stops are likely to

---

[6] The expected F0 changes (Hz) depending on speaker's age (in months) should be yielded as follows:

F0 change in lenis stops = $364.48 + (-1.75\times\text{Age})$
F0 change in fortis stops = $31.06 + 364.48 + (0.091 + (-1.75))\times\text{Age}$
F0 change in aspirated stops = $-30.23 + 364.48 + (1.88 + (-1.75))\times\text{Age}$

share the overlapping F0 values, but afterwards children come to distinguish aspirated stops with the highest F0 values compared to fortis and lenis stops.

The lack of statistical significance of the relationship between age and fortis stops leads us to consider the possibility that F0 is not a critical acoustic parameter to phonetically (and phonologically) characterize fortis stops, so the accuracy level in F0 would not contribute to the production of fortis stops even by the older children. This possibility is in line with the finding from Kong *et al.*'s study (2011) that only VOT plays a significantly important role in defining fortis stops in production. The significantly short VOT already contributes enough information to phonetically distinguish fortis stops; the other acoustic parameter of F0 mainly functions to distinguish between the other two stop categories, lenis and aspirated stops.

The next model was designed to find the correlation between gender and stop categories. Through this model, we are able to see how significantly *Gender* affects phonetic differentiation in F0 between three different stop groups. In spite of the fact that no significant gender effect on VOT was found in the three stop categories, the question of how gender could affect the general F0 developmental pattern with respect to distinction of stop contrasts can still be considered. The results are illustrated as follows.

$$(6') \; F0_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Male_i + \beta_3 \times Fortis_{ij} + \beta_4 \times Aspirated_{ij} + \beta_5 \times Fortis_{ij} \times Male_i + \beta_6 \times Aspirated_{ij} \times Male_i + \eta_i + \varepsilon_{ij}$$

(Reference categories: *Lenis* for *Stop category, Female* for *Gender*)

Table 3.24. The output of the mixed-effects linear model based on (6′).

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	945.3	30.75	
Residual		5105.9	71.46	
Number of obs: 2331, groups: childid, 58				
Fixed effects:				
	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	350.7492	23.2963	15.056	< 2e-16 ***
Fortis	31.9617	5.1872	6.162	8.82e-10 ***
Asp	43.3191	5.2097	8.315	2.22e-16 ***
Age	-1.1206	0.5999	-1.868	0.0670 .
Male	-23.5147	10.2809	-2.287	0.0244 *
Fortis:Male	6.8642	8.4335	0.814	0.4158
Asp:Male	-11.877	8.3612	-1.42	0.1556

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The results show that in the lenis stop category, female child speakers tend to have significantly higher F0 values than male child speakers (*coef.* = -23.5147, $p < 0.01$). Mainly because of the physical condition of the vocal cords, female adult speakers usually have much higher F0 values than male adult speakers, but in child speech this gender effect was not expected. In fortis and aspirated stops, no significant gender difference in F0 is likely to be found. Therefore, through this model, it can be said that gender effects on F0 distinction of stop contrasts are not clearly observed.

As shown previously in the control group, POA obviously affects VOT in the same stop category, and it has been questioned whether POA could affect F0 values. Interestingly, when the same child data was analyzed in terms of VOT, an apparently significant correlation between POA and VOT variation was not found, which implies that the child speakers have not yet learned VOT perfectly, even though their stop category differentiation using VOT was quite successful. This time the effect of POA on F0 ranges was investigated with the following statistical model in (7′), which includes the interaction term *Stop category* × *POA*.

$$(7') \; F0_{ij} = \beta_0 + \beta_1 \times Fortis_{ij} + \beta_2 \times Asp_{ij} + \beta_3 \times Alveolar_{ij} + \beta_4 \times Velar_{ij}$$
$$+ \beta_5 \times Fortis_{ij} \times Alveolar_{ij} + \beta_6 \times Fortis_{ij} \times Velar_{ij} + \beta_7 \times Asp_{ij} \times Alveolar_{ij}$$
$$+ \beta_8 \times Asp_{ij} \times Velar_{ij} + \eta_i + \varepsilon_{ij}$$

(Reference categories: *Lenis* for *Stop category*, *Labial* for *POA*)

Table 3.25. The output of the mixed-effects linear model based on (7′).

Random effects:

Groups	Name	Variance	Std. Dev.
childid	(Intercept)	952.6	30.86
Residual		5100.6	71.42

Number of obs: 2331, groups: childid, 58

Fixed effects:

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	353.8617	24.3865	14.511	< 2e-16 ***
Alveolar	-1.0326	7.4335	-0.139	0.88957
Velar	5.5248	6.7696	0.816	0.41454
Fortis	37.0342	6.2871	5.891	4.56e-09 ***
Asp	40.005	6.521	6.135	1.04e-09 ***
Alveolar:Fortis	-15.5744	11.1023	-1.403	0.16084
Velar:Fortis	-1.0462	9.207	-0.114	0.90954
Alveolar:Asp	-4.8897	10.2044	-0.479	0.63187
Velar:Asp	0.771	9.551	0.081	0.93567

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

In general, the predicted fixed effects of the predictors are not significant. In particular, it cannot be said that the correlation between *Stop category* and *POA* is statistically significant and it is also notable that the interaction terms have high standard errors. This might be because the analyzed data consists of a number of tokens that is not great enough to divide into six different categories. Another possible cause could be that the variation in F0 is not unified in any way simply because there is no physiological motivation for POA to differentiate F0 values.

3.7. Discussion

The findings from the production experiment with the toddler speakers consistently indicate that the development of VOT and F0 is necessarily linked to the phonetic distinction of Korean stop contrasts. Regarding aspects of the developmental pattern of VOT or F0, children's age is directly related. Across the three different phonation types, older children were more likely to produce significant phonetic differences than younger children. This pattern seems clear and apparent, but within the same age range widely different phonetic values were estimated as well.

This study attempted to discover more generalized developmental trajectories of VOT or F0 in relation to the production of Korean stop categories, which consist of three different phonation types, using more specified statistical methods. Interestingly, the three different statistical methods have reported slightly different results for significance of the correlation between children's age and VOT or F0.

As expected, VOT is directly relevant in phonetically characterizing fortis stops, and it appears that the acquisition of VOT occurs prior to F0, referring to the results of the mixed-effects models, which showed that fortis stops have significantly shorter VOT values and the correlation between age and VOT in fortis stops is the strongest. These findings imply that for the accurate acoustic implementation of fortis stops, the role of VOT is determinant, and that during the target ages (2 to 4 years), significant VOT development occurs. On the other hand, VOT values in lenis and aspirated stops did not show unified patterns, which indicates that along with the role of VOT in the production of lenis and aspirated stops, an F0 distinction is required. This implication is not in line with the report from Kong *et al.* (2011) that the perception of aspirated vs. lenis productions of young children also depends on the role of VOT. On the other hand, this finding is in support for the suggestion from Kang (2014) that a great deal of overlap in VOT between lenis and aspirated stops has been commonly found in adult speech in recent Seoul Korean. Thus, the most accurate production of fortis stops by effective VOT differentiation by young children suggests a universal acquisition pattern that voiceless unaspirated stops are acquired earlier than the other two phonation types (Kim, 2008).

In terms of F0, more complicated patterns were observed. It seems that child speakers understand the role of F0 as a useful tool to show the phonetic difference between aspirated and lenis stops, since the simple regression showed that age and F0 differences between those two stop types were the most significant. More specifically, the mixed-effects models showed that there was not a significant correlation between F0 in fortis stops and age, while lenis and aspirated stops were likely to have some significant interactions with children’s age. This can be an indication of the weakest relationship between fortis stops and F0. The actual F0 values that were phonetically estimated were not unified in general, and in the case of lenis stop production, the values had large standard errors. At the same time, there was a significant correlation between age and F0 in lenis stops. This fact indicates that the role of F0 is crucial for the phonetic differentiation of lenis stops, and the acquisition of lenis stops would be the most difficult, and would therefore come latest, for children compared to other stop manners. The difficulties in acquisition of the lenis stops by young children would be accounted for by physiological reasons and phonetic complexity of F0. Lenis stops have long lag VOT values but they still have lowered F0 values. It would be difficult to make laryngeal adjustment to lower F0 for AP-initial lenis stops for young children (Kim, 2008). Lenis stops are realized as voiced intervocalically (e.g., /t/ → [d]/ V_V) and distinct F0 variations for phrase medial lenis stops are hardly observed due to overall phrasal intonation (Cho et al, 2002; Jun, 1996), so it is understandable that acquisition of lenis stops can be challenging for young toddlers compared to the other two phonation types.

As the adult control group showed that the three different stop manners have significantly distinctive F0 and VOT values, the child target group's articulatory distinction of Korean stop types still needs to develop further to be accurate and specific. The statistical analysis implies that there is significant development of the acoustic implementation of the three phonation types in relation to those acoustic parameters between 2 years and 4 years of age. It seems that the acquisition of fortis stops would have developed the most along with VOT over this age range, and that aspirated stops would be acquired next due to the contrastive acoustic feature with the longest VOT. Still, more accurate differentiation of F0 remains for the children to acquire in the process of the acquisition of lenis and aspirated stops.

Chapter 4
Perception experiments

In order to understand children's perceptual ability, I first set up various perception tests that seemed suitable for toddlers; through experience with the target-aged toddlers, I decided to use an identification test. It was not easy to find an optimal and appropriate method to investigate their current learning stages. The target ages are 2 to 4 years, so I had to make the test suitable for toddlers with a 2-year age gap. During the period of 2 to 4 years of age, an enormous amount of language development occurs, so the same task would simultaneously be hard for the younger children but easy for the older children. Therefore, this study focused on balancing the degree of difficulty of the task and confirming that all the target-aged children were able to do the task.

Before conducting an actual experiment, I designed a pilot discrimination test in which children would decide whether the given sounds were the same or different. In this case, relatively younger children did not understand what the word 'same' or 'different' meant, and sometimes they did not know what 'sound' meant. Secondly, I asked them to label the given sounds. Most of the children were capable of finishing this task, but it requires their concentration for a while since the task must continue for more than 18 trials (9 pairs of stop sounds × 2 switched trials, in addition to filler trials). To make the experiment effective and efficient, I decided to shorten and simplify the task so that the probability of success would increase and toddlers would not get easily distracted. The methods used will be explained in detail in the following section, 4.1.

If the formation of a perception system precedes the development of production, perceptual ability should be superior to articulatory skills. As previously suggested in this study, young Korean children tend to utilize VOT as one of the important acoustic parameters for producing appropriate contrasts in the Korean stop system at an early age, around 2 years (e.g., Y.-T. Kim, 1996; M. J. Kim & Pae, 2005; Kong, 2009). If their stop productions with VOT were successful, then it is naturally expected that their perceptual distinction with VOT must be more developed. As shown in the previous production experiments, phonetic differentiation with VOT, especially for Korean fortis stops, was significantly successful in general. This fact implies that perceptual discrimination using VOT changes could be much more successful at an earlier age. The child participants' ages are all over 2 years, so it seems unnecessary to determine whether they could successfully perceive their native phonetic features with VOT differences in Korean stop contrasts. Therefore, one of the primary goals to pursue is to examine the development of F0, which is considered to be acquired later than VOT, through the interaction between perceptual representation and acoustic implementation. As many studies have suggested, perceptual sensitivity of young children decreases as they get the input of their native language resulting in the formation of their phonetic categories for native contrasts (e.g., Kuhl *et al.*, 1992; Polka & Werker, 1994; Werker & Tees, 1984). In Korean, F0 makes a phonological stop contrast in word-initial position; so the relevant perceptual representations for the three different stops in the dimension of F0 should undergo phonetic adjustment until children complete the acquisition of stop contrasts. Focusing on these properties, this study aims to reveal the learning process according to children's age through a perception test.

In this study, I designed a perception experiment in order to investigate how young Korean children developed phonemic processing through adjusting their perceptual sensitivity to F0, and how closely a certain age and the perceptual categorization of Korean stop manners are correlated. In this perception experiment, I hypothesize that a perceptual F0 threshold for a certain Korean stop category has been developed and stabilized with age. I assume that the observed perceptual threshold for a stop category is also responsible for F0 distinction in production of aspirated and lenis stops recalling that VOT can exclusively characterize fortis stops in production.

This study will test this hypothesis by investigating the developmental process of phonetic categorization for lenis and aspirated stops in the dimension of F0 on a perceptual map. Specifically, stop sound tokens with a certain range of F0 changes at the vowel onset right after the target stop sound were manipulated manually, and the synthesized sound tokens were prepared for 48 young Korean children to identify those stimuli. In the experiment, the stimuli were manipulated to determine whether Korean toddler speakers are able to discriminate an F0 continuum covering a range from 200 Hz to 295 Hz.

In this chapter, two different kinds of perception experiments are introduced. The first was designed in order to grasp a general perception pattern regarding Korean stop contrast, focusing on determining which stop categories would be most and least correctly perceived. Through this process, I could establish mastery ordering among the three different stop categories in relation to children's age. The main assumption of the experiment is that fortis stops are acquired earlier than lenis and aspirated stops due to their distinctive VOT values. The second experiment aims to reveal the correlation between the developmental trajectory of F0 and the perception of lenis and aspirated stops. I track the change of children's perceptual sensitivity to F0 and its consequences for the perceptual distinction of lenis-aspirated stop contrast.

The children's responses were analyzed using mixed-effects logistic regression modeling. Through this statistical method, the correlation between children's responses and predictor variables such as their age and phonetic differences can be investigated.

4.1. Perception experiment 1

This section deals with the perception experiment that was performed in order to describe the general perceptual ability of young Korean child speakers regarding three different Korean stop manners. Recalling the results of the previous production experiment in which the young children showed relatively mature productions of fortis stops with distinctively short VOT values, it is expected that the perception of fortis stops would be more accurate and successful than that of lenis or aspirated stops.

4.1.1. Participants

Of the 58 children who participated in the production experiment, 48 children, aged between 2;0 and 3;11 (years;months), also participated in the perception experiment. There were no reported hearing or speaking disorders, and the children's parents or guardians were all Korean-language monolinguals. The child participants are divided into four different age groups, as shown in the table below.

Table 4.1. Child participant information for a perception test.

Age group (years;months)	Male	Female	Total
2;0-2;6	5	6	11
2;7-3;0	4	4	8
3;1-3;6	8	7	15
3;7-3;11	3	11	14
Total	20	28	48

4.1.2. Procedure

The identification test was performed with a point-to-a picture task. The 18 trials consisted of nine minimal pairs, which included every possible pair of lenis-fortis-aspirated homorganic stops. Using triplets in a point-to-a-picture task was avoided since it would make the test harder; choosing one of three alternatives might be too difficult for young children at the target age. In addition, finding appropriate triplets that young children would know would be almost impossible. Therefore, I chose to use three minimal pairs with two alternatives. The list of words used is presented in the table below.

Table 4.2. The pairs and words used in the perception test.

POA & Stop category	Pairs	Word	English gloss	Word	English gloss
Labial lenis-fortis	/p/-/p'/	/paŋ/	room	/p'aŋ/	bread
Labial lenis-aspirated	/p/-/p^h/	/pal/	foot	/p^{h}al/	arm
Labial fortis-aspirated	/p'/-/p^h/	/p'ʊl/	horn	/p^{h}ʊl/	grass
Alveolar lenis-fortis	/t/-/t'/	/tal/	the moon	/t'al/	daughter
Alveolar lenis-aspirated	/t/-/t^h/	/tɔk'i/	hammer	/t^{h}ɔk'i/	rabbit
Alveolar fortis-aspirated	/t'/-/t^h/	/t'ʌk/	rice cake	/t^{h}ʌk/	chin
Velar lenis-fortis	/k/-/k'/	/kʊl/	cave	/k'ʊl/	honey
Velar lenis-aspirated	/k/-/k^h/	/kɔŋ/	ball	/k^{h}ɔŋ/	beans
Velar fortis-aspirated	/k'/-/k^h/	/k'inʊn/	wearing	/k^{h}inʊn/	turning on

Each pair was presented twice in random order. Before beginning the experiment, each participant played with the experimenter using the pictures (and other toys) so the participant was aware of what each picture depicted naturally during playtime. In addition, all the pictures had already been shown to them in the previously conducted production experiment, so it was assumed that all child participants knew the name for each picture.

For the identification test, the target sounds were 18 different stop sounds, but the two trials that were prepared for identifying aspirated stops from the lenis-aspirated pairs of pictures (i.e., identification of /p^{h}al/ and /k^{h}ɔŋ/) were excluded from analysis in this section, since those pairs were analyzed in another perception experiment with synthesized sounds, which will be explained in detail in the following section 4.2.2.[7] In the identical pairs of lenis-aspirated stops, natural productions of lenis stops were

---

[7] The results of the identification of the pair of /p/-/p^h/ and /k/-/k^h/ were dealt with in the second perception experiment with synthesized sound tokens (of /pal/ and /kɔŋ/). The two different kinds of stimuli (natural and synthesized sounds) were tested in the same session. The two aspirated stops were provided as synthesized, so the two trials for natural /p^{h}al/ and /k^{h}ɔŋ/ were excluded in the session. This is to prevent the children having certain phonetic standards for natural /p^{h}al/ and /k^{h}ɔŋ/, which can interfere with the perception of synthesized counterparts. As a result, the identification of aspirated stops in the natural productions of /p^{h}al/ and /k^{h}ɔŋ/ was excluded in this section, and only natural productions of /pal/ and /kɔŋ/ were included for analysis.

prepared and provided during this experiment.

After the children heard the stimuli, which were recorded by a female adult speaker, they were asked to choose or point to one of the paired pictures. Their responses were documented during the session by the experimenter.

4.1.3. Results

During the session, 768 responses (16 trials × 48 child participants) were obtained. Generally the results show that almost all child participants successfully identified the given natural sounds.

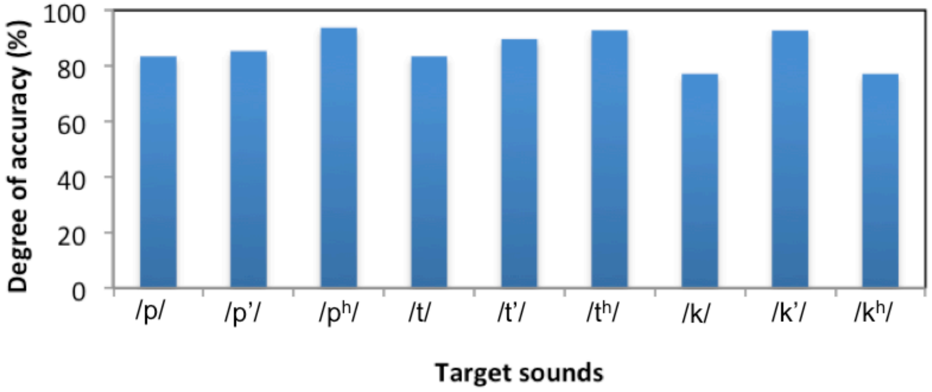

Figure 4.1. The degree of accuracy for each target sound in the identification test.

As Figure 4.1 illustrates, almost every natural sound stimulus was correctly identified (average 86.7%). The lowest accuracy rate was found in the identification of velar lenis /k/ (77.1%) and the most successful identification was of labial aspirated /p^h/ (93.8%). At a glance, it can be observed that the children's perceptual identification was relatively challenging in the case of velar stops. Figure 4.1 effectively sums up the overall results, but it does not show the actual differences between trials.

The accuracy levels of each trial are presented as follows. Table 4.3 shows which trials were easiest and most challenging for the children. The accuracy levels for the identification of the same target phoneme differ depending on what minimal pairs were provided for the choice. For example, alveolar lenis stop /t/ was correctly identified 91.7% of the time when it was provided with the alternative of /t'/. However, when the children had /t/ and /t^h/ as their choices, only 75% of the responses were correct in the identification of /t/. It is hard to generalize about perception by language-learning children, since it seems that in some ways their perceptual bias from the given context and the lexicon affects their responses more than fine phonetic details. Therefore, it is also important to understand what the children did in each trial.

Table 4.3. The degree of accuracy of each trial in the perception test with natural stimuli.

Answer	Used pairs as alternatives	Degree of accuracy (%)
p	/p/-/p'/	89.6
p'	/p/-/p'/	91.7
p	/p/-/p^h/	77.1
p'	/p'/-/p^h/	83.3
p^h	/p'/-/p^h/	93.8
t	/t/-/t'/	91.7
t'	/t/-/t'/	83.3
t	/t/-/t^h/	75.0
t^h	/t/-/t^h/	93.8
t'	/t'/-/t^h/	95.8
t^h	/t'/-/t^h/	95.8
k	/k/-/k'/	87.5
k'	/k/-/k'/	100.0
k	/k/-/k^h/	66.7
k^h	/k'/-/k^h/	77.1
k'	/k'/-/k^h/	85.4

The toddlers seemed to be easily confused when the two alternatives had a large F0 difference rather than a VOT difference. For instance, when the toddlers were provided with velar lenis /k/ as a target sound and they had to select their response between /k/ and velar aspirated /k^h/, their accuracy rate was the lowest (66.7%). The most apparent phonetic difference between lenis and aspirated stops is an F0 difference. On the other hand, relatively large VOT differences allowed them to correctly identify the stimuli. For example, velar fortis /k'/ was the easiest phoneme to identify when it was provided with lenis stop /k/ (100%). This tendency was also found in the case of identification of /p/. It seemed easier for children to identify /p/ when it was paired with /p'/ (89.6%), while the same stop /p/ was not well identified when with /p^h/ (77.1%).

This nonuniform pattern in the children's responses, which depends on the amplified phonetic difference, implies that the child speakers' perceptual ability can vary depending on the given context. The results in the table above represent the children's responses overall, while there was a big difference found in the accuracy levels between relatively younger and older children. Therefore, a detailed explanation of the children's responses will be explored considering the weights of the age difference and what the given alternatives were.

4.1.3.1. Lenis and fortis stops

This section is devoted to comparing accuracy levels in the cases when lenis and fortis stops were provided in the same trial. The words used consisted of the lenis-fortis minimal pairs in Figure 4.2, given as paired alternatives for children to select. Recalling the fact that the production of fortis stops by the same toddlers is successful relative to lenis or aspirated stops, it is expected that the perceptual identification of fortis stops would also be more successful. Indeed, the target sounds were correctly perceived by the children at a rate of almost 90%. The velar fortis /k'/ was most successfully identified (100%), while its lenis counterpart /k/ was correctly identified at 87.5%. Generally, it seems natural that the identification of fortis stops is much more successful, but in the case of alveolar stops, /t/ is correctly perceived much more frequently compared to fortis /t'/. Younger children tend to be biased toward more familiar words, so there is a possibility that their vocabulary affects their accuracy levels. For example, the word /tal/ ('moon'), which was used for the test, is likely a relatively familiar word to most of the young children in Korea. The other alternative /t'al/ ('daughter') that was used for a selective question could be a more challenging word to learn for younger children because of its gender-related meaning, compared to the gender neutral word, /tal/. For the word pairs, it could be possible that an imbalance in frequency of occurrence or word comprehension affects the responses in the experiment. However, this does not necessarily invalidates the experiment, since in the previous production experiment, which was conducted in the same session, the same pictures were used in order to minimize the risk of this possibility.[8]

---

[8] According to McArthur Communicative Development Inventory-Korean (Pae *et al.*, 2004) and Shin (2005), lenis stops occur most frequently word-initially in children's words while fortis stops occur least frequently in that position, and there was no evident relation of phoneme frequency to early acquisition.

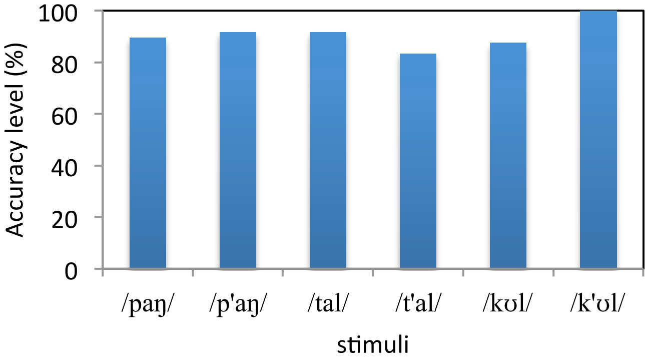

Figure 4.2. Accuracy levels according to target words in lenis-fortis minimal pairs.

The response patterns by four different age groups are shown in Figure 4.3. The youngest group shows the lowest accuracy levels in three different stop categories overall, and the oldest group shows the highest accuracy levels except in the case of /kʊl/. The youngest age group was more successful in the task when the target sounds were fortis stops rather than lenis stops. For example, while the youngest group perceived /k’/ most accurately, the alveolar lenis /t/ was least accurately perceived. This fact suggests that perception of the two different stop categories does not develop simultaneously and that the perception of fortis stops, which heavily depends on distinctive VOT values, develops earlier than that of lenis stops.

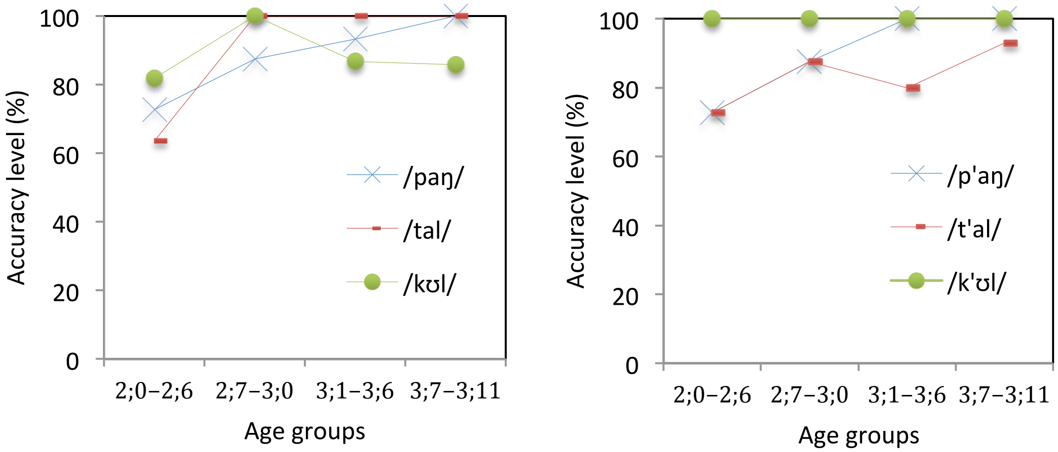

Figure 4.3. Accuracy level by age group for lenis (left panel) and fortis stops (right panel).

It is also notable that the perceptual patterns of these two stop categories vary depending on what words are used, and that the toddlers' response patterns are not affected solely by the phonation type itself. Even though the oldest group showed the lowest accuracy level in the perception of /kʊl/, we cannot exclude the possibility that the children's lexicon affects their responses rather than a delay in their acquisition of /k/ occurred compared to /p/ or /t/, in which their perceptions are 100% accurate.

4.1.3.2. Fortis and aspirated stops

This section deals with the children's perception of fortis and aspirated stops when they were minimal pair alternatives. The two stop categories are considered to have a subtle F0 difference while the VOT difference is considerably large. Because of the extreme difference in VOT between fortis and aspirated stops, it is expected that the children's responses would be mostly correct when the words used consisted of fortis-aspirated minimal pairs. If we assume that the children have developed VOT as an important acoustic cue for Korean stop distinction, the words with a word-initial fortis stop should be correctly perceived.

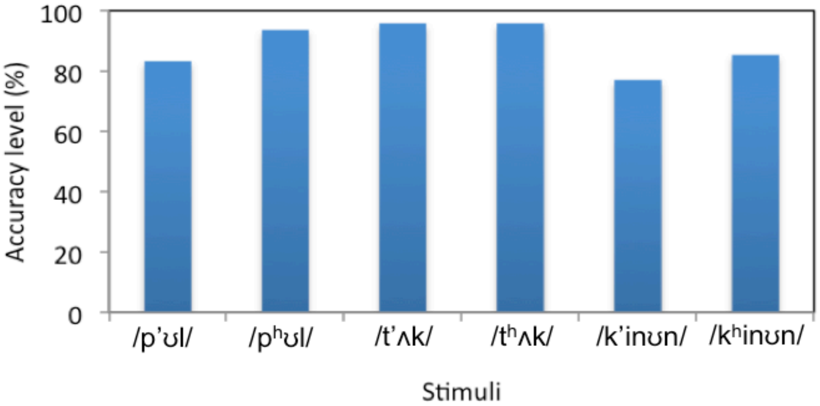

Figure 4.4. Accuracy levels according to target words for fortis-aspirated minimal pairs.

Perceptual identification by the children is similarly successful compared to the case of lenis-fortis minimal pairs. Considering that both the fortis-aspirated and lenis-fortis sets have a large VOT difference between minimal pairs, it is plausible that the children were able to correctly identify stop type when VOT difference was emphasized. In general, it seems that fortis stops did not tend to be correctly perceived as much as aspirated stops. Rather, in the labial and velar pairs, the perception of aspirated stops shows higher accuracy levels. The fortis velar stop /k'/ was the most difficult sound for the children to perceived correctly (77.1%). Perception of the alveolar pair seems most successful. The words that start with /t'/-/t^h/ were correctly identified at the same accuracy rate (95.8%).

As a result, superior perception of fortis stops was not found in any of the three different minimal pairs.

The following figure presents how children's age affects their responses.

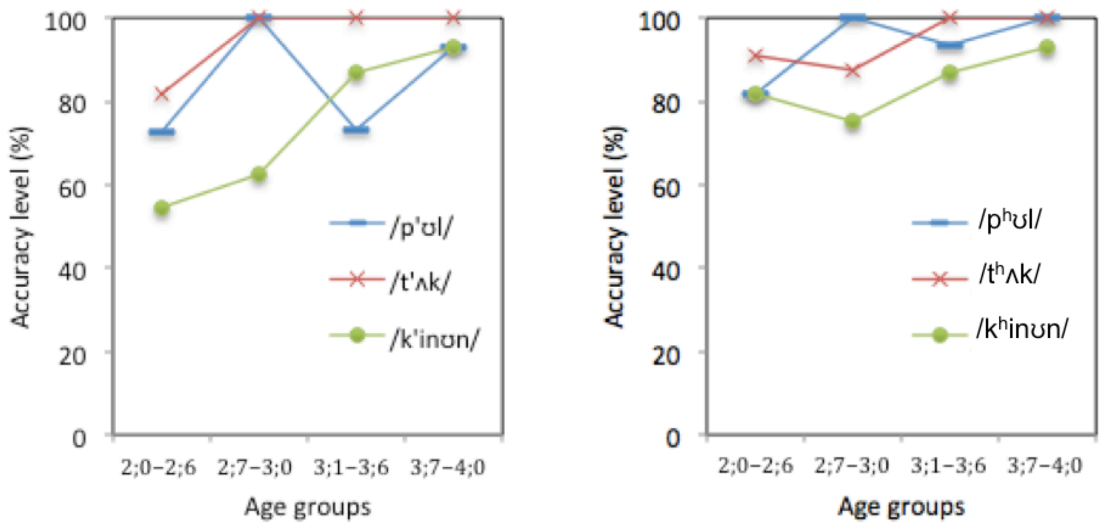

Figure 4.5. Accuracy level by age group for fortis (left panel) and aspirated stops (right panel).

All of the child participants identified aspirated stops more successfully (average 91.7%) than fortis stops (average 85.4%), except for the perception of the alveolar set by the second youngest group, which showed the same accuracy levels in both stop categories (average 87.5%). Based on these results, the perception of fortis stops could be confusing to the toddlers, and VOT development may not be solely related to the perception of fortis stops. It seems that younger children's perception depends partly on their lexicon or the context, since aspirated counterparts were correctly identified, in contrast with the fortis stops, in most cases. Therefore, it can be suggested that lower scores in the perception of fortis stops do not directly indicate that VOT development has not occurred or is delayed.

As seen in Figure 4.5, the youngest age group generally showed the least successful results, but the other three age groups also showed a variety of perception patterns. Compared to the perceptual patterns for aspirated stops, the accuracy levels in the perception of fortis stops vary, which indicates that children's perceptual accuracy is not directly or absolutely dependent on their age. The labial pair /p'/-/p^h/ were successfully identified by the second youngest group (both 100%), while the velar pair /k'/-/k^h/ was most correctly perceived by the third youngest group (both 86.7%).

In the case of aspirated counterparts, perceptual accuracy is generally stable since accuracy levels for the three words with aspirated onsets are over 70% in all cases. However, among the fortis counterparts, the three different words show quite different accuracy levels. It is a surprising finding that the accuracy levels for labial fortis /p'/ do not gradually increase according to age because it is usually expected that labial fortis stops are relatively easily learned compared to alveolar or velar stops. In spite of the various perceptual patterns in the younger age groups, the oldest age group shows that their perception of the two different stop categories is quite stable, indicating that before 4 years of age, acoustic parameters to distinguish the two stop categories have been acquired and operate quite adequately.

4.1.3.3. Aspirated and lenis stops

Due to the fact that the two aspirated stop trials with natural sounds /p^{h}al/ and /k^{h}ɔŋ/ were not included in the test, I now have only four different trials consisting of three lenis stops and one aspirated stop (alveolar /t^h/) in this section. The pair of aspirated and lenis stops has large F0 differences but relatively subtle VOT differences. Considering that F0 development would occur later than VOT development, perception in this case was expected to be most difficult for the children.

The results in Figure 4.6 revealed that the perception of the aspirated stop is more accurate (93.8%) than the perception of the other three words with lenis onsets. The least accurate identification was of the word with an initial velar lenis stop /kɔŋ/ (66.7%). The identification of /pal/ and /tɔk'i/ was relatively successful and the accuracy levels were similar (77.1% and 75% respectively).

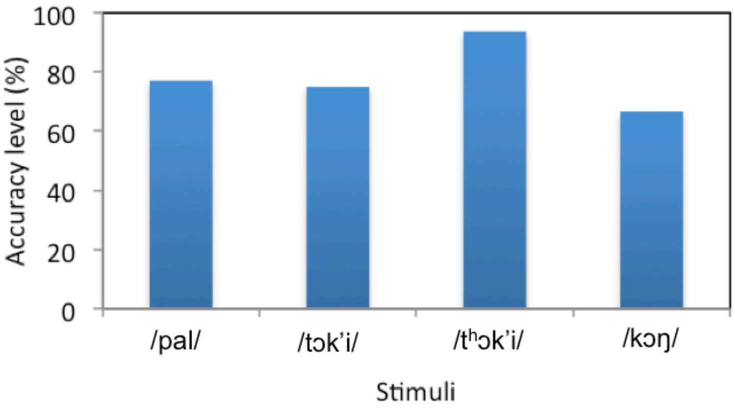

Figure 4.6. Accuracy levels according to the target words for lenis-aspirated minimal pairs.

The difference in accuracy levels depending on age groups is presented in Figure 4.7. Generally, perception of aspirated stop /t^h/ is much better than others even though the scores did not absolutely increase with age; the third age group showed the lowest score (87%) while the second youngest and the oldest group showed perfect identification (100%). Meanwhile, the three lenis stops were not correctly identified in many cases across all the age groups. It is notable that the perceptual identification of lenis stops showed the least successful result when they were matched with aspirated counterparts. However, recalling that the same three lenis stops tended to be identified correctly more when they were paired with fortis counterparts as alternatives, it appears that fortis stops can be relatively more perceptually salient when VOT difference is more significant and amplified; lenis stops were relatively easily differentiated by VOT differences.

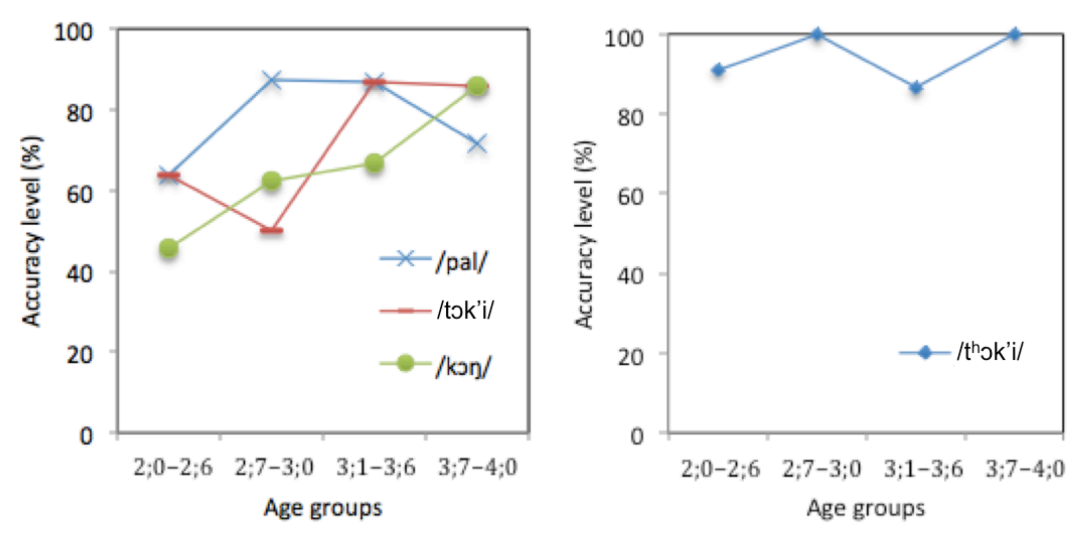

Figure 4.7. Accuracy level by age group for lenis (left panel) and aspirated stops (right panel).

The perception of aspirated stops was relatively successful even by the youngest group, while the perceptual development of lenis stops is more directly related to age. The better perception of aspirated stops might indicate that aspirated stops are more perceptually salient than their lenis counterparts. This superior perceptual identification seems understandable with the general typology of stops with VOT ranges; voiceless aspirated stops can be phonetically identified by their long VOT range (Keating, 1985; Lisker & Abramson, 1964). As previously noted, typologically common stop systems usually include voiceless aspirated stops, whereas lenis stops are typologically rare due to their laryngeal features (Cho & Ladefoged, 1999; Maddieson, 1984). Because of this distinctive phonetic difference, it is likely that the perceptual development pertaining to aspirated stops would occur earlier than for lenis stops.

However, I have to point out that the two stop groups mostly share similar VOT ranges, resulting in dependence on F0 difference for distinction. Therefore, it can be assumed that higher F0 ranges for aspirated stops are more perceptually salient than lower F0 ranges for lenis stops. Another possibility to account for the perceptual salience of aspirated stops is related to the acoustic characteristics that are produced through a phonetic combination of the longest VOT values and the highest F0 values. In Kong *et al.* (2011), VOT is also responsible for the identification of aspirated stops in the case of adult speakers' perception of Korean stop contrasts. Additionally, it is also notable that lenis stops are voiced when they occur intervocalically, so VOT ranges of lenis stops can differ depending on the position, while aspirated stops always have the same phonetic realizations in terms of voicing (Kochetov & Kang, 2016). This allophonic variation of lenis stops might affect the acquisition of lenis stops by young children. Thus, the perceptual identification of lenis stops could be much more challenging compare to the perception of aspirated stops in spite of the fact that perception of both of the two stop categories is directly related to F0 dimensional distinction.

The figure above represents a consistent phenomenon throughout the analysis in that the accuracy levels in the perceptual identification of lenis stops vary depending on what the target word was. This non-uniform pattern itself implies that F0 development is still in progress, so the children are not able to identify the lenis stops with perfection especially when provided with aspirated counterparts. The youngest age group showed unsuccessful perception – that is, around or under 60% accuracy. The oldest age group showed 86% and 71% of perceptual accuracy, so it can be said that perceptual identification of lenis stops seems generally challenging for the young children, especially when the target words were of lenis-aspirated minimal pairs.

4.1.4. Statistical analysis

Through the results of the perception experiment, there seems to be a relationship between age and perceptual accuracy in spite of the observed differences in the perceptual identification of the three different phonation types. The accuracy levels for each stop category are as follows.

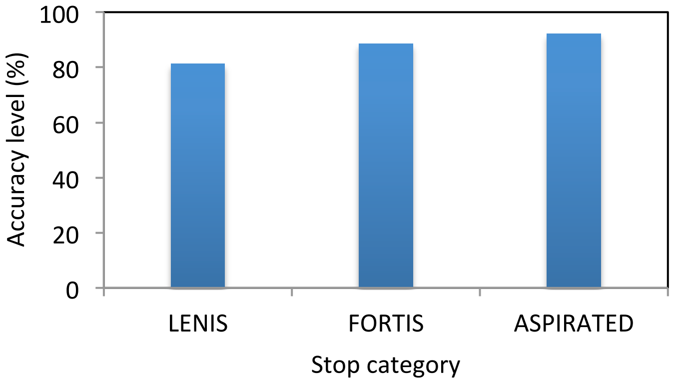

Figure 4.8. Accuracy levels for three different stop categories in the perceptual identification task with natural stimuli.

The three different phonation types were generally well-identified by the child participants, though not perfectly, since accuracy levels are all over 80%. This perceptual pattern indicates that the children's acoustic space with multiple parameters regarding stop contrast has developed for perceptual discrimination while an accuracy level of 80% means that phonemic perception with fine phonetic details is not yet perfect. As mentioned earlier, there are other factors that matter to children's perception such as lexical familiarity, context, and attention span. Children's lexicon plays an important role in their perception, especially at an early age, since if their acoustic development has not yet completed, they need to depend on the contextual information available. During this experiment, this kind of effect was restricted so that children had to determine their answers relying mainly on the perceived acoustic differences. Considering this possibility, it can be concluded that their acoustic space for correct phonemic mapping will develop further.

It is quite surprising that perceptual accuracy for fortis stops is not the highest all the time, since fortis stops have been considered to be the most perceptually salient compared to lenis or aspirated stops due to their distinctive VOT values (Kong *et al.*, 2011). As an acoustic cue, VOT has been acquired earlier, so it is expected that fortis stops that are more directly related to VOT should be more successfully perceived. However, the poorer result in fortis stops does not necessarily indicate that VOT has not been acquired or developed. In fact, aspirated stops display the longest VOT values, so it seems that the perception of aspirated stops is also linked to VOT. The main reason for the highest perceptual accuracy of a specific phonation type is its distinctive VOT values. Aspirated stops also have the highest F0 values compared to lenis or fortis stops, so the phonetic combination of VOT and F0 could make them more perceptually salient. However, as Kong (2009) noted, adult speakers' perception of aspirated stops also depends on VOT and F0 values (even though there is a tendency that female adult speakers depend more on perceived F0 differences). Assuming that this would apply to the perception of the child participants, it seems more convincing that the perception of aspirated stops is more salient because of their distinctively long VOT values and high F0 values. If we assume that the acoustic parameter of VOT has been fully developed during 2 to 4 years of age, while F0 still has to develop more after this stage, then the crucial role of VOT can also account for this perceptual pattern. Because lenis stops have intermediate VOT values, which should be less salient in perception, it would be the main reason for the lowest accuracy level in this test (e.g., Cho *et al.*, 2002; Y.-T. Kim, 1965; M. Kim, 2008).

The following figure shows the accuracy-level differences by different age groups. As expected, the accuracy levels in the perceptual identification of aspirated and fortis stops similarly increase according to age. In the case of lenis stops, accuracy levels are the lowest across all age groups, which implies that the children cannot perceptually distinguish the acoustic characteristics of lenis stops compared to the other two series of stops.

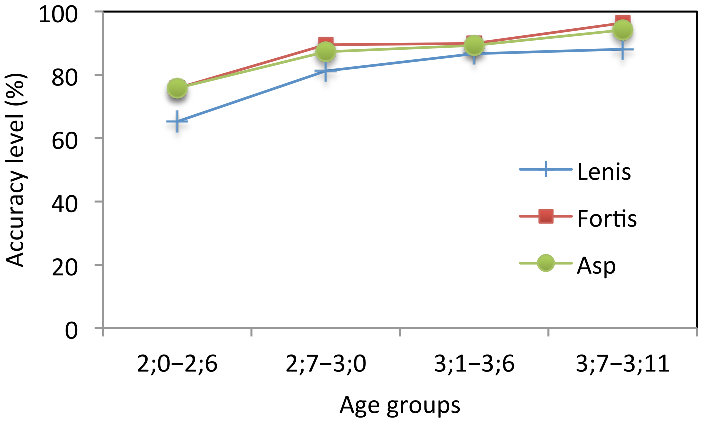

Figure 4.9. Accuracy level differences by age group in three stop categories.

So far, the results give a strong impression that lenis stops are the most delayed in perceptual acquisition, while fortis and aspirated stops are acquired relatively early due to their perceptual salience. This perception pattern may be led by VOT values of fortis or aspirated stops since VOT is a more easily acquired acoustic parameter compared to F0.

In order to confirm the effect of predictor variables on the observed perceptual accuracy of the toddlers, a mixed-effects logistic regression model (Gelman & Hill, 2006;

Raundenbush & Bryk, 2002; Snijder & Bosker, 1999) was designed and conducted. The *lme4* package (Bates *et al.*, 2015) was implemented in R, and *glmer* function was used. The following model intended to show the effect of *Age* on the accurate perception of the three phonation types. *Lenis* was the reference category.

(8) $$log\left(\frac{(Answer_{ij})}{1-(Answer_{ij})}\right) = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Fortis_{ij} + \beta_3 \times Aspirated_{ij} + \eta_i + \varepsilon_{ij}$$

(When *Answer* is correct, *Answer* = 1, otherwise 0)

Table 4.4. The output of the mixed-effects logistic regression model based on (8).

Random effects:

Groups	Name	Variance	Std. Dev.
childid	(Intercept)	2.652e-17	5.15e-09

Number of obs: 768, groups: childid, 48

Fixed effects:

	Estimate	Std. Error	z value	Pr (> |t|)
(Intercept)	-1.99584	0.687	-3.299	0.000971 ***
Age	0.09538	0.01686	5.656	1.31e-08 ***
Fortis	0.63412	0.24462	2.592	0.009535 **
Asp	1.07337	0.31442	3.414	0.000641 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

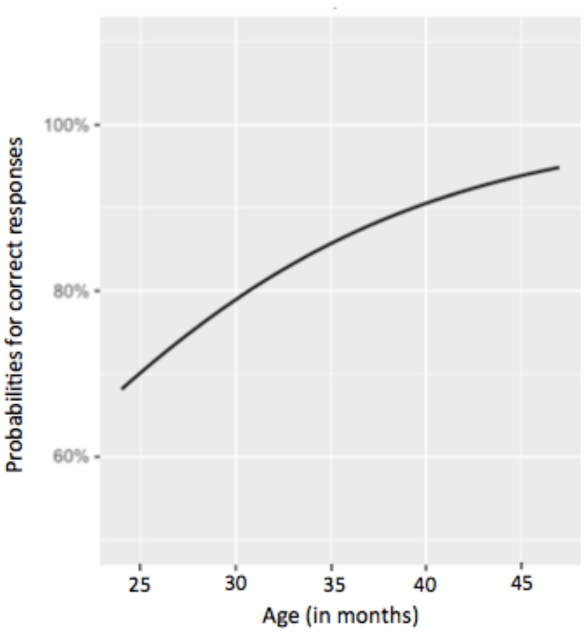

Figure 4.10. The relationship between age and correct answers with a logistic curve for all responses by the child participants.

Significant effects of *Age*, *Lenis*, *Fortis*, and *Aspirated stops* are observed through this statistical model (*p*-values < 0.05). The fixed-effects predictors, *Age*, *Fortis*, and *Aspirated,* have positive coefficients, which means that these predictors are positively related to the probabilities of correct answers in the perception experiment: the children's perceptual accuracy increases with age, and perception of fortis and aspirated stops is more accurate than that of lenis stops. However, this model only predicts the effect of each variable if it independently affects the responses.

The following model was used to predict the interaction of *Stop category* and *Age* in the children's responses. *Lenis* was the reference category.

$$(9)\ log\left(\frac{(Answer_{ij})}{1-(Answer_{ij})}\right) = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Fortis_{ij} + \beta_3 \times Aspirated_{ij} + \beta_4 \times Fortis_{ij} \times Age_i + \beta_5 \times Aspirated_{ij} \times Age_i + \eta_i + \varepsilon_{ij}$$

Table 4.5. The output of the mixed-effects logistic regression model based on (9).

Random effects:

Groups	Name	Variance	Std. Dev.
childid	(Intercept)	3.294e-13	5.74e-07

Number of obs: 768, groups: childid, 48

Fixed effects:

	Estimate	Std. Error	z value	Pr (> |t|)
(Intercept)	-1.47039	0.83135	-1.769	0.07695 .
Age	0.08044	0.02321	3.465	0.00053 ***
Fortis	-0.65035	1.31834	-0.493	0.62180
Asp	0.46016	1.66768	0.276	0.78260
Age:Fortis	0.03802	0.03815	0.997	0.31899
Age:Asp	0.01807	0.04835	0.374	0.70857

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The fixed-effects coefficients show that most of the predictors, except for *Age* ($p < 0.001$), do not have significant effects on correct identification by the children (p-values > 0.05). Here, no significant correlation between children’s age and the perceptual accuracy of fortis or aspirated stops can be found ($p > 0.05$), but there is a significant interaction between age and the successful perception of lenis stops ($= -1.47039 + (08044 \times age)$). That is, with increase in age (in months), there is an 8% increase (*coef.*≈ 0.08) in the odds of correct answers for lenis stops. This could be interpreted as a significant development of the related acoustic parameters during the period of 2 to 4 years of age. A lack of significant interaction between age and the correct identification of fortis stops indicates that during the same period, there seems to be no significant change in children’s acoustic space with respect to VOT. Alternatively, as this statistical model is tracking only the change itself depending on each predictor, the lack of significance for predictors during the age of 2 to 4 years might be interpreted as no significant change during that period, in which even the younger children successfully perceived those stop categories to the same extent as the older children.

The next model in (10) was built to determine whether there would be a significant POA effect in perceptual identification by the children. The results showed no significant effect of a specific POA on perceptual accuracy in the identification of stop contrasts.

(10) $$log\left(\frac{(Answer_{ij})}{1-(Answer_{ij})}\right) = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Alveolar_{ij} + \beta_3 \times Velar_{ij} + \beta_4 \times Fortis_{ij} + \beta_5 \times Aspirated_{ij} + \beta_6 \times Fortis_{ij} \times Age_i + \beta_7 \times Aspirated_{ij} \times Age_i + \varepsilon_{ij} + \eta_i$$

Table 4.6. The output of the mixed-effects logistic regression model based on (10).

Random effects:

Groups	Name	Variance	Std. Dev.
childid	(Intercept)	3.025e-13	5.5e-07

Number of obs: 768, groups: childid, 48
Fixed effects:

	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-1.40093	0.87544	-1.651	0.098659 .
Age	0.08069	0.02325	3.473	0.000514 ***
Fortis	-0.6491	1.32046	-0.492	0.623023
Asp	0.46529	1.66904	0.279	0.780414
Alveolar	0.09412	0.27877	0.338	0.735651
Velar	-0.32113	0.26838	-1.231	0.218291
Age:Fortis	0.03747	0.03801	0.986	0.324292
Age:Asp	0.01747	0.04812	0.362	0.716556

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

4.1.5. Discussion

Through this perception experiment with natural sound stimuli, we can conclude that perceptual accuracy in the identification of lenis stops is the lowest, and that in general, this perceptual pattern is similarly applicable to the four different age groups. It also seems that young children aged 2 to 4 years could correctly perceive fortis and aspirated stops compared to lenis stops. It is assumed that the perceptual system surrounding those phonation types has not been perfectly stabilized as much as adult speakers' perception, since their perceptual accuracy was not close to 100%.

However, these imperfect perceptual patterns do not necessarily mean that their perceptual system with regard to the three-way stop contrast has not developed enough to facilitate linguistic development. The results should also be considered with the fact that the toddlers had to identify the stimuli in an experimental setting, in which contextual information about the target phonemes was extremely restricted. In the actual session, it could be possible that they depended on the lexical familiarity or other lexical properties when the perceived acoustic difference was not apparent to them. Another possible cause for lowered perceptual accuracy is toddlers' short attention span. The session consisted of simple serial trials, which might cause the children to get distracted and bored.

Despite the fact that these kinds of interfering factors in the analysis may make the results seem inconclusive, the toddlers' responses still suggest that their perceptual mapping functions distinctively based on the VOT dimension rather than the F0 dimension. Referring to the production data that was gathered from the same toddler speakers and the control group, it was revealed that fortis stops were quite exclusively distinctive with their VOT values, so it can be suggested that the acoustic dimension of VOT has developed enough to more accurately perceive fortis stops. Because lenis stops have intermediate VOT values, and possibly a VOT merger with aspirated stops, it seems that the perceptual identification of lenis stops was not an easy task for the children. In addition, if we hypothesize that their acoustic space (in terms of F0) has not developed enough to distinguish phonetic difference in F0 between lenis and aspirated stops, the lowest perceptual accuracy in the identification of lenis stops can be explained. As we recall from the logistic mixed-effects analysis, only the perception of lenis stops showed a significant change in perceptual accuracy with age. This indicates that during the period of 2 to 4 years of age, there would occur a significant linguistic development regarding the perception of lenis stops, which implies the phonemic development in the F0 dimension.

The following figures show how the toddler speakers responded to three different lenis stops and how each age group responded to lenis stops with fortis or aspirated counterparts. When lenis stops were paired with fortis counterparts in minimal pairs, the listeners must have focused on VOT differences between the two phonation types. On the other hand, when lenis stops were paired with aspirated stops in minimal pairs, the F0 difference was amplified due to VOT merger. Considering this comparison, it is plausible that VOT differences would be more perceptually salient than F0 differences across three different POAs to the toddlers as shown in Figure 4.11. The velar lenis /k/ was the least perfectly identified by the toddlers (66.7%) when it was compared to aspirated /k^h/. The perception of lenis stops was the most difficult task for the toddlers, but with fortis counterparts lenis stops tended to be successfully identified. This means that in children's perception systems, the same phoneme can be perceived differently depending on what phonetic parameter is dominantly involved in the perception.

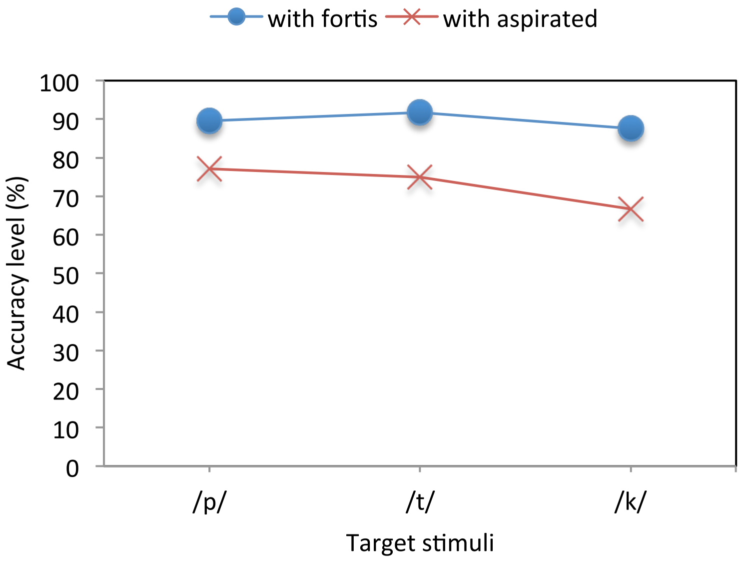

Figure 4.11. Accuracy levels in the perceptual identification of lenis stops by POA, when provided with fortis or aspirated minimal pairs.

Figure 4.12 illustrates change in perceptual accuracy observed in perception of lenis stops according to age. When VOT differences are maximized with fortis counterparts, lenis stops were almost perfectly identified even by the second age group (95.8%). However, with a maximal F0 difference from the aspirated counterparts, the oldest age group only showed an 81% accuracy level in the perception of lenis stops. This means that phonemic categorization in the F0 dimension still needs to develop further at 4 years of age, while the VOT dimension already functions for the appropriate phonemic categorization. This fact is also confirmed in Figure 4.13, which compares overall accuracy levels across the perception of the three different stop types. When VOT differences were amplified more than F0 differences, the toddlers correctly perceived the stimuli at a rate of 90%. However, when they had to concentrate on F0 differences between the minimal pairs, only 78.2% of the responses were correct.

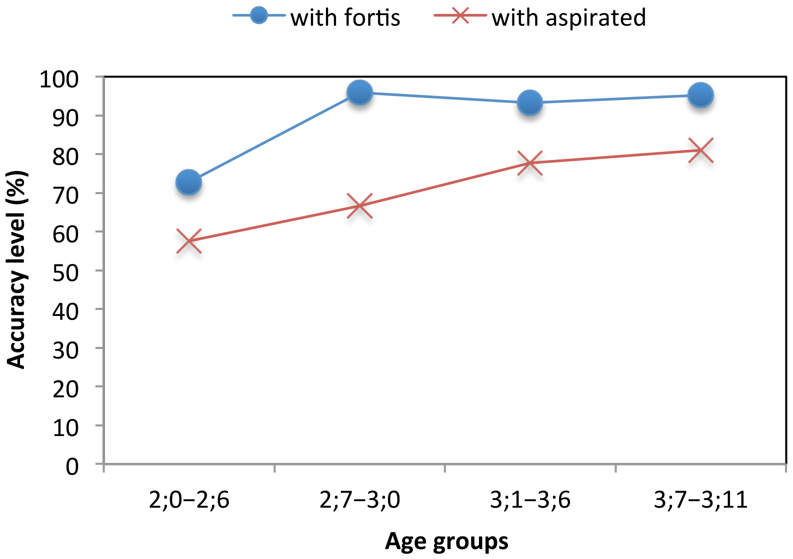

Figure 4.12. Comparison of accuracy levels in the perceptual identification of lenis stops by age group, when provided with fortis and aspirated minimal pairs.

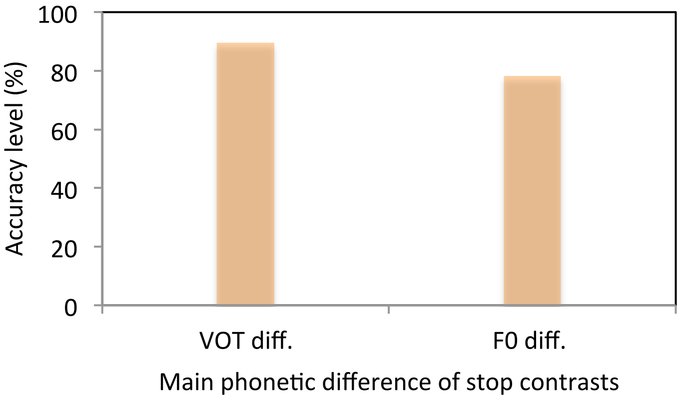

Figure 4.13. Accuracy level differences for children's perception of VOT difference and F0 difference between the stop contrast.

Therefore, the results of the perception test with natural sound stimuli imply that, before the age of 2 years, children's perception system for Korean stop contrasts has started to develop in the VOT dimension and to operate to distinguish stop types, since perception of fortis and aspirated stops was relatively successful. The lowest perceptual accuracy with lenis stops indicates that lenis stops need another acoustic dimension to be distinguished in their perceptual space, and that F0 has not yet developed enough to allow perfect phonemic categorization of lenis stops during the period of 2 to 4 years of age.

4.2. Perception experiment 2

Perceptual sensitivity is adaptive to a native contrast, so it decreases with age resulting in categorical perception (Kuhl, 2004; Werker, 1995; Werker & Pegg, 1992). For categorical perception, perceptual space with relevant native acoustic parameters should develop and function for native phonemic categories. For Korean stop contrasts, the acoustic dimensions of VOT and F0 are necessary to process phonemic categorization. The development of these two acoustic parameters promotes this organizing process, and it is directly related to children's linguistic competence. As we recall that fortis stops are correctly perceived even at an early age, and that VOT seems to be almost exclusively related to the perception of fortis stops, phonetic distinction between lenis and aspirated stops requires the F0 dimensional distinction. The need for a robust F0 difference between lenis and aspirated stops in adult speech has been found in many studies (Choi, 2002; Kang, 2014; Kang & Guion, 2008; Kim, 2004; Kong *et al.*, 2011; Lee & Jongman, 2012; Oh, 2011; Silva, 2006; Wright, 2007). However, it is still questionable whether the children's perception system already recognizes F0 differences categorically as adult speakers do. It has been suggested that children over 2 years of age perceive their native phonetic details in phonemic speech perception (Werker & Baldwin, 1992; Werker & Pegg, 1992). Assuming that F0 necessarily develops in phonemic processing for Korean stop contrasts, this perception experiment hypothesizes that during the period of 2 to 4 years of age, a significant F0 development occurs. Since it provides a significant distinction between aspirated and lenis stops in children's perception system, this experiment attempts to show how an F0 threshold for aspirated stops changes according to age. This analysis suggests a developmental trajectory of F0 as an acoustic parameter to distinguish contrastive phonemes with the lenis-aspirated stop contrast.

4.2.1. Participants

The same child participants who participated in the previous perception test were the subjects of this perception test.

4.2.2. Procedure

One pair of pictures at a time was presented to each child participant using a personal laptop. The participant was asked to point to one of the two given pictures right after hearing "*mwueti* [target word] *ici?*" meaning 'which one is [target word]?' from the audio. The children were encouraged to touch the laptop screen, and if they touched one of the minimal pair pictures when they heard the stimulus, the pair of pictures for the next trial would pop up. The stimuli that were used in this point-to-a-picture task were synthesized stimuli. They were mixed in random order and provided to children in the test. The perception experiment was conducted directly after the production experiment finished, in the same room. The session took approximately 10 minutes and consisted of 37 trials (5 fillers + 16 synthesized stimuli + 16 natural stimuli).[9] If a child seemed distracted, the same question was asked again. Every answer by the participant was documented by the experimenter during the test.

The productions of /pal/ ('foot') and /kɔŋ/ ('ball') by a female speaker were used for the synthesis. Using the pitch synthesis function in Praat (Boersma & Weenink,

---

[9] The two different kinds of stimuli were mixed in random order and used in the same session for the identification task. The main reason for mixing those stimuli was that toddlers' concentration would not be maintained for two different sessions. The two different stimuli need fillers anyway, so mixing together was the best option, as they acted as each other's natural fillers. The five fillers that were used only as fillers consisted of bi- or tri-syllable words. The 16 natural stimuli are separately dealt with in Section 4.1.

2015), the onset of the following vowel was manipulated to have a 15 Hz difference between the tokens. The vowel onset of /pal/ was manipulated to have 200 Hz, 215 Hz, 230 Hz, 245 Hz, 260 Hz, and 275 Hz with fixed VOT (70 ms). The vowel onset of /kɔŋ/ was manipulated to 220 Hz, 235 Hz, 250 Hz, 265 Hz, 280 Hz, and 295 Hz with fixed VOT (75 ms). During synthesis, VOT was fixed so that the six different tokens had the same VOT values. As F0 values at vowel onset were changed, the consecutive pitch points were also changed with the same difference so that the original pitch contour was nearly maintained, resulting in stimuli that sound natural at its maximum.

Table 4.7. Synthesized stimuli.

/pal/	F0 (Hz)					
	200	215	230	245	260	275
/kɔŋ/	F0 (Hz)					
	220	235	250	265	280	295

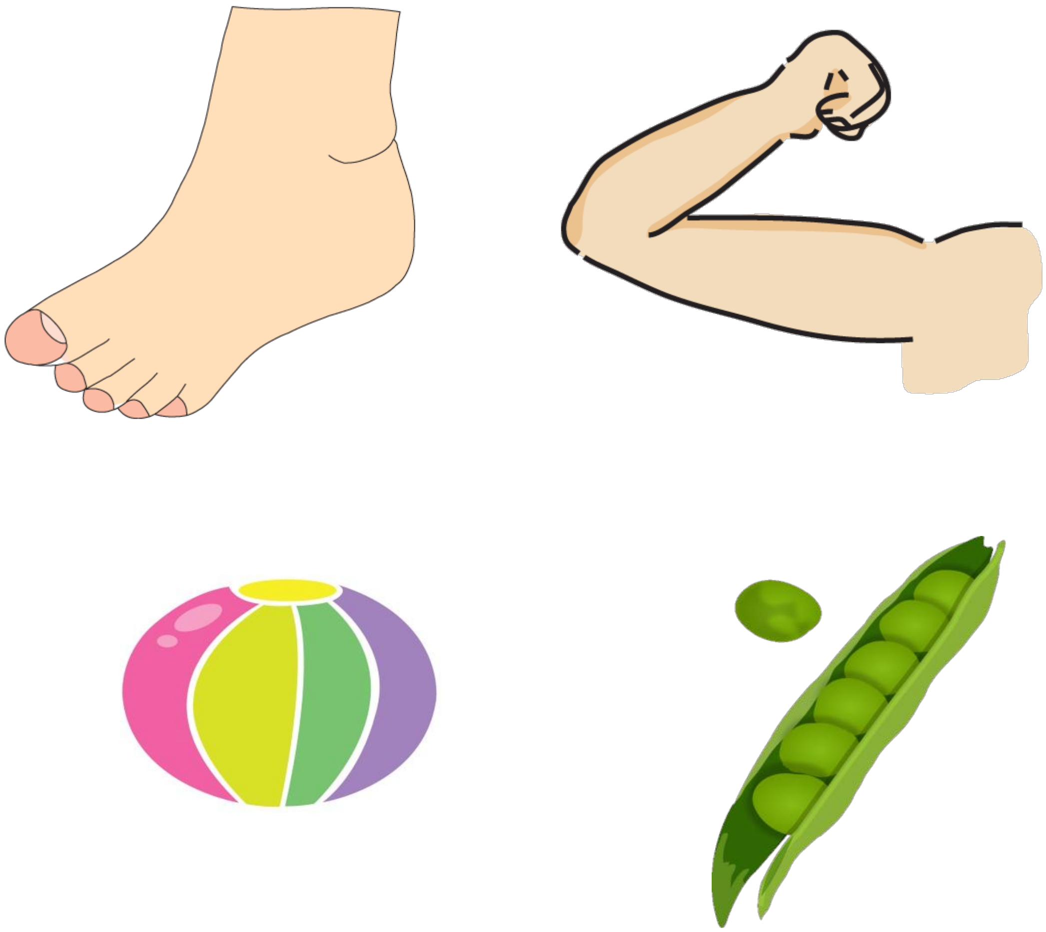

Figure 4.14. Two paired sets in the perception test. /pal/-/pʰal/ (above) and /kɔŋ/-/kʰɔŋ/ (below).

4.2.3. Results

The synthesized set of /pal/-/pʰal/ shows that the children's lenis-aspirated stop distinction relies on a substantial F0 difference between lenis /p/ and aspirated /pʰ/. The lenis /p/ with 200 Hz at its vowel onset, which is considered of an average lenis articulation, was identified as /p/. On the other hand, the synthesized /pal/ with 275 Hz at its vowel onset was perceived as aspirated /p^h/ in 80% of the responses. Figure 4.15 shows how the child speakers perceived each synthesized stimulus.

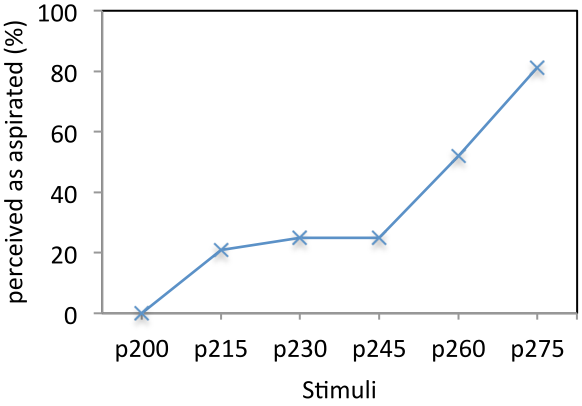

Figure 4.15. Results of the identification of synthesized /p/. Six stimuli with a 15-Hz difference are presented on the x-axis. The numbers following ‘p’ indicate manipulated F0 values at vowel onset.

Perceptual identification relates to F0 values of the words that have a synthesized word-initial stop, even though their perception does not seem to be categorical. The synthesized /p/ stimuli with 15 Hz, 30 Hz, and 45 Hz difference tended to be perceived as lenis /p/, but an F0 difference of 60 Hz or more meant that the stimuli were more likely to be identified as the aspirated counterpart /p^h/. Thus, F0 values at vowel onset affect children’s perceptual identification patterns, which indicates the effect of F0 on perceptual discrimination between lenis and aspirated stops. Of course, there seems to be a need for a substantial F0 difference to some extent in order for the lenis stop to be perceived as aspirated; a subtle change in F0 difference does not necessarily produce a perceptual distinction by young children. For example, a 15 Hz to 45 Hz difference would not make the stimuli be perceived as aspirated in many cases, which is shown by the fact that only 20% of the stimuli were identified as aspirated /p^h/, as seen in Figure 4.15. In order to be identified as aspirated /p^h/ in at least half of the responses, an F0 difference around 52.5 Hz (= (the midpoint of 245 and 260) – (the standard value of the lenis /p/ = 200 Hz)) seems to be required.

Not surprisingly, all of the children did not provide stable responses, meaning that their responses are not consistent with regard to the stimuli. Their perceptual sensitivity to F0 differences was not consistent. For example, one child identified a stimulus with 215 Hz as aspirated /p^h/, but identified the stimulus with 245 Hz as lenis /p/. Each stimulus was tested once, so it was impossible to calculate the average response per stimulus. At this point, I needed some standard to determine which responses could be considered stable and consistent and thus apt for analysis, and which ones should be excluded from the analysis due to inconsistency. To make judgments about the goodness of the response stability, a simple logistic regression model was conducted for a series of responses of each child participant (in which *Aspirated* =1 and *Lenis* =0). The results suggested that over 10.0 of AIC would not be acceptable as a stable response, so only the responses with a number under 10.0 of AIC were used for analysis.[10] Therefore, in the current study, if the child's confusion persists over two steps of trials, I concluded not to take it as a valid response. If their choice was reversed within two steps, it was

---

[10] AIC is 4.0, but a perfect reversal as 1 0 0 0 0 0 for 220 Hz, 235 Hz, 250 Hz, 265 Hz, 280 Hz, and 295 Hz, respectively, was excluded (only two cases in the /k/ set: *childid_13*, *childid_22*).

considered a stable response.[11] Table 4.8 and Figure 4.16 show the number of children in each age group who showed response stability.

Table 4.8. The number of subjects and stable responders for /p/ in four different age groups.

Age groups	2;0-2;6	2;7-3;0	3;1-3;6	3;7-3;11
Total number of subjects	11	8	15	14
Number of stable responders	7	7	13	14
% of stable response	63.6%	87.5%	86.7%	100%

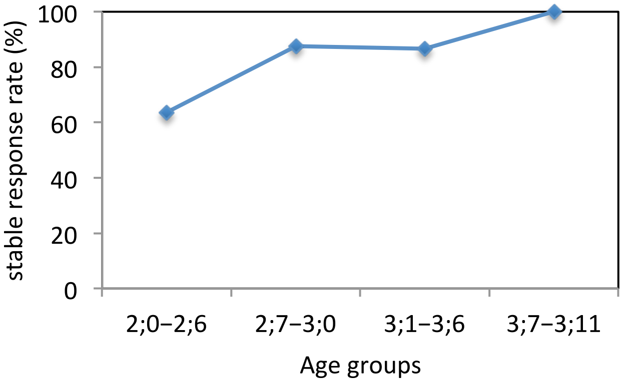

Figure 4.16. The ratio of stable responses to all obtained responses to synthesized /p/.

In general, the stable response rate gradually increases according to age. This pattern indicates that the children's phonetic standards for determining aspirated stops in their acoustic perceptual map have stabilized to develop according to age, which makes them able to more consistently perceive aspirated stops using relative F0 differences. This also

---

[11] If a child responded 0 0 0 1 0 1 for 200 Hz-215 Hz-230 Hz-245 Hz-260 Hz-275 Hz, respectively, this was a stable response. If a child chose 0 0 1 0 0 1, this was not a valid response. In the latter case, there are two lenis choices between the two aspirated choices. It is considered unstable because it can interfere with the reliability of the analysis as follows, so I decided not to use this kind of inconsistent response.

means their perceptual sensitivity to F0 decreases as they grow up. This process promotes categorical perception depending on their native phonological contrasts. The oldest age group shows 100% stable responses, so it can be suggested that before 3;6 of age, the perceptual threshold for aspirated /p^h/ uniformly and consistently operates in the distinction between labial lenis and aspirated stops.

Surprisingly, the results of the velar stop contrast /k/-/k^h/ do not correspond to those of their labial counterparts. Figure 4.17 shows how the same children responded to the velar stimuli. Overall, the responses do not change accordingly depending on the relative difference in F0 for /k/-/k^h/ distinction. The perceptual pattern is hardly generalized with irregular curves, and a large amount of phonetic difference in F0 did not guarantee a higher percentage of words to be perceived as aspirated. The only case in which over 50% of the cases were identified as aspirated /k^h/ is the stimulus with 295 Hz. This is an outstanding difference from the labial contrast, since 52.5 Hz difference seemed to be required in order for the stop to be perceived as /p^h/, while /k/ requires around 67.5 Hz (= 287.5 Hz (= the midpoint between 295 Hz and 280 Hz) − 220 Hz (= the plain /k/)) difference for it to be perceived as /k^h/.

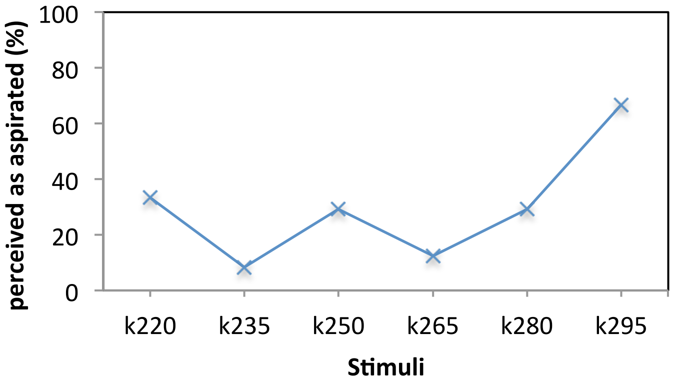

Figure 4.17. Results of the identification of synthesized /k/. Six stimuli with a 15-Hz difference are presented on the x-axis. The numbers following ‘k’ denote manipulated F0 values at vowel onset.

These irregular perception patterns reveal that the children had a hard time in the perception of velar stops /k/ and /k^h/ compared to labial stops /p/ and /p^h/. Difficulties in identifying lenis or aspirated in velar stops could be caused by intrinsic acoustic characteristics of velar stops such as stronger aspiration than labial stops. Compared to bilabial stops, velar stops are characterized by their stronger aspiration in articulation, and this phonetic information is crucial when identifying /k^h/ as well. Considering this intrinsic acoustic feature, the role of VOT in velar stop perception seems a reasonable explanation, but it remains questionable why in adult speech there is a very small amount of VOT difference between the productions of lenis and aspirated stops, as Kang (2014) recently showed VOT merger between lenis and aspirated stops in adult speech in Seoul Korean. In addition, the speech of the control group in the previous production experiment also showed that the degree of VOT difference is similar across stop categories with different POAs (mean VOT difference: labial=20.636 ms, velar=15.322 ms).

Another possibility to explain the poorer perception of velar stops is lexical differences in toddlers' lexicons. There is a possibility that the words in the minimal pair, /kɔŋ/ ('ball') and /k^{h}ɔŋ/ ('beans'), could not be equally familiar enough to distinguish in an experimental setting. There is a strong tendency for toddler speakers to choose a more familiar word, especially when they might not be sure about what they heard. On the other hand, this poorer perceptual pattern for /k^h/ and /k/ could be caused simply by the possibility that the acquisition of velar stop contrast may be delayed compared to labial stop contrast. With the limitations of the tested stimuli, the only thing to surely note is that a relative F0 difference around 67.5 Hz is required to perceptually identify the velar stop /k^h/ from the lenis /k/. It is a surprising finding that the velar aspirated stop /k^h/ requires a much larger F0 difference from the lenis stops compared to the labial stop contrast. That is, the same amount of F0 difference cannot guarantee the same degree of perceptual distinctiveness between lenis and aspirated stops.

In spite of the fact that the velar stop pair showed poorer and less consistent perceptual patterns, the response stability grew according to the participants' ages, which is represented in Figure 4.18. The first and the second youngest age groups show very similar response stability (45.5% and 37.5%, respectively), while there is a relatively big gap in response stability between the second and third age groups (37.5% vs. 73.3%). In the third age group, 73% of the toddler participants aged 3;1 to 3;6 consistently responded to the stimuli. Less than half of the toddlers in the first two age groups did not respond consistently, which indicates that even 3-year-old toddlers did not have fixed stable criteria in the F0 dimension to distinguish lenis and aspirated stops. The oldest age group showed more consistent responses (78.6%) than younger groups, which is similarly observed in the case of the synthesized /p/. However, the oldest age group's responses were not perfectly consistent.

Table 4.9. The number of subjects and stable responders for /k/ in four different age groups.

	2;0-2;6	2;7-3;0	3;1-3;6	3;7-3;11
Total number of subjects	11	8	15	14
Number of stable responders	5	3	11	11
% of stable response	45.5%	37.5%	73.3%	78.6%

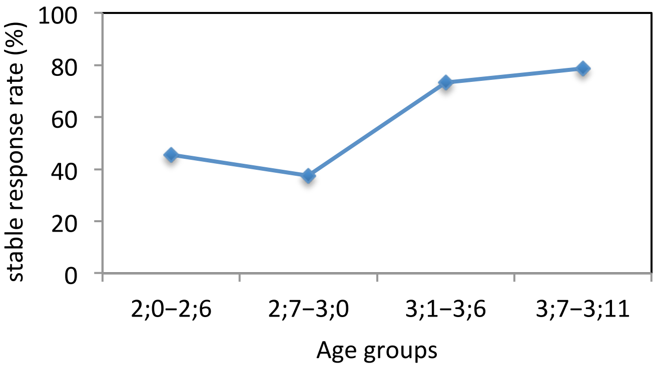

Figure 4.18. The ratio of stable responses to all obtained responses to synthesized /k/ by age group.

Recalling the response stability in the case of synthesized /p/, the oldest age group's responses were 100% consistent, but there was a relatively sharp increase in response stability between the first and second youngest age groups. This means that at an age between 2;7 and 3;0, toddlers have experienced critical development of F0 that operates for perceptual distinction between labial lenis and aspirated stops. Along the same lines, the velar stop distinction develops between the age of 2;7 and 3;6, which is delayed a little compared to the labial stop distinction.

These observations help to convey a clear picture of the F0 developmental trajectory, which indicates that F0 development, which is directly involved in perceptual discrimination between lenis and aspirated stops, occurs around 3 years of age, while F0 plays a critical role for the labial stop distinction at an earlier age compared to the velar stop distinction.

4.2.4. Statistical analysis

As the dependent variable has binary values, a logistic regression model was constructed and applied. First, in order to determine a significant relationship between the toddlers' responses and their age or stimuli, the mixed-effects logistic model in (11) was conducted. The predictor variable *F0_difference* means the amount of relative F0 difference each stimulus has from the lenis stimulus, so 6 different values in a Hz-scale (i.e., 0 to 75, which means 0 Hz to 75 Hz difference from the lenis stop) were used in this continuous variable. The model in (11) was applied for the responses to the synthesized /p/.

(11) $log\left(\frac{(Answer_{ij})}{1-(Answer_{ij})}\right) = \beta_0 + \beta_1 \times Age_i + \beta_2 \times F0\_difference_{ij} + \varepsilon_{ij} + \eta_i$

(*Answer* = 1 if a child's response was aspirated stops, otherwise *Answer* =0)

Table 4.10. The output of the mixed effects logistic regression model for synthesized /p/.

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	1.553	1.246	
Number of obs: 288, groups: childid, 48				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-3.111203	1.461000	-2.130	0.0332 *
Age	-0.016345	0.036946	-0.442	0.6582
F0_difference	0.067825	0.009396	7.219	5.26e-13 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

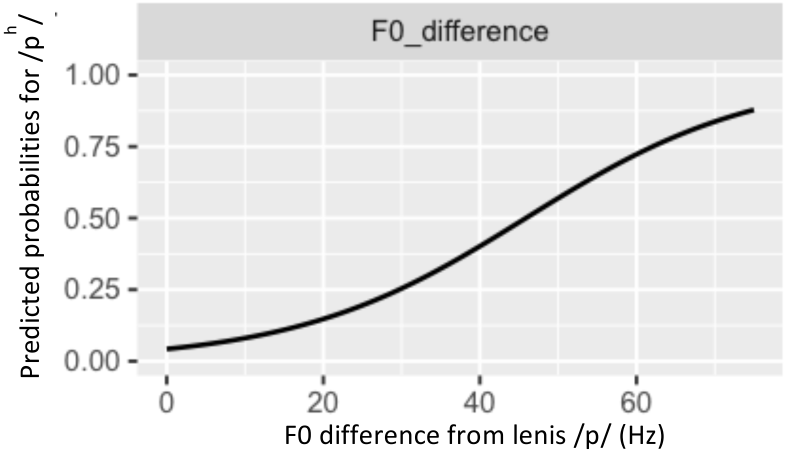

Figure 4.19. Relationship between F0 differences and the perception of aspirated /p^h/ with a logistic curve.

It was quite expected that the relative F0 difference affects the children’s responses, and this model confirms the significance of F0 difference in the perception of aspirated /p^h/. The fixed-effects coefficient for *F0_difference* is 0.0678. This means that there is a 7% increase ($1.07 \approx e^{0.0678}$) by F0 difference in the odds of the probability of the stimuli being identified as aspirated /p^h/. However, the toddlers' age has no significant effect on their responses. An insignificant relationship between toddlers' age and their responses seems natural in this model because older/younger age cannot guarantee a higher/lower probability of perceiving aspirated stops. Therefore, another model (12) with an interaction term was built and conducted in order to investigate the effects of age on the responses for each stimulus.

(12) $$log\left(\frac{(Answer_{ij})}{1-(Answer_{ij})}\right) = \beta_0 + \beta_1 \times Age_i + \beta_2 \times F0\_difference_{ij} + \beta_3 \times Age_i \times F0\_difference_{ij} + \varepsilon_{ij} + \eta_i$$

(*Answer* = 1 if a child's response was aspirated stops, otherwise *Answer* =0)

Table 4.11. The output of the mixed-effects logistic regression model for interaction effect of *Age* and *F0_ difference* in the case of synthesized /p/.

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	1.841	1.357	
Number of obs: 288, groups: childid, 48				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	3.067115	2.414631	1.27	0.20401
Age	-0.194295	0.070236	-2.766	0.00567 **
F0_difference	-0.066039	0.041896	-1.576	0.11497
Age:F0_difference	0.00381	0.001233	3.09	0.00200 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

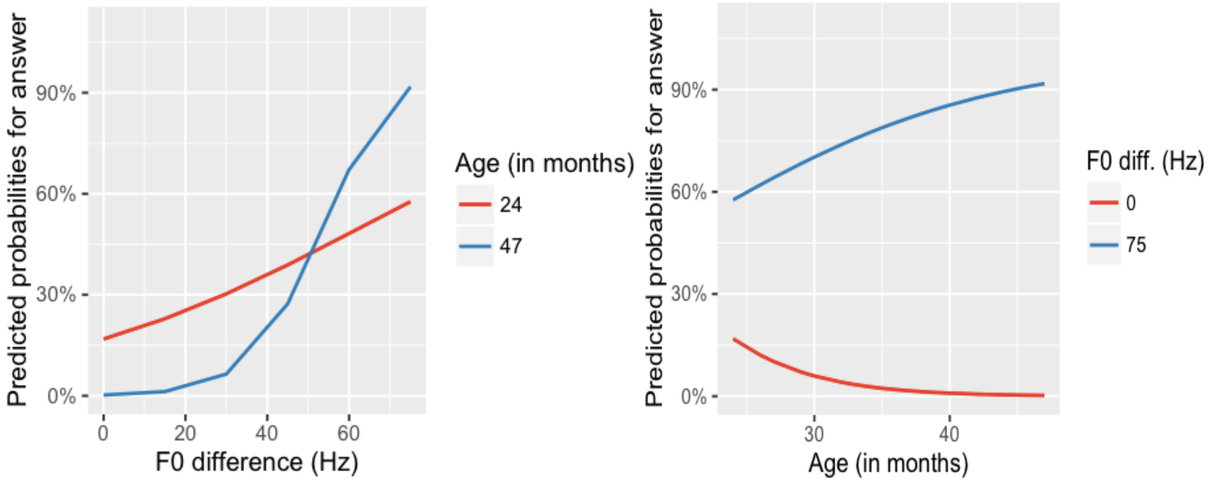

Figure 4.20. Interaction effect of *Age* and *F0_difference* on *Answer* of /p^h/.

The fixed-effects coefficient for the interaction between *Age* and *F0_difference* shows that it significantly affects the children's responses (*coef.* = 0.00381, $p < 0.01$). For each stimulus with a different F0 value, the probability of perceiving it as aspirated /p^h/ is significantly related to the children's age. The associated figures above show clearly that if a stimulus has a bigger F0 difference from lenis /p/, it is more likely to be perceived as /p^h/. In particular, this tendency is mostly found in the case of relatively older children, so the significant effect of the interaction between *Age* and *F0_difference* on the toddlers' responses seems assured. This interpretation also recalls the higher response stability in the oldest age group. In the case of synthesized /p/, the children's perceptual patterns were nearly consistent according to the stimuli with different F0 values. The output of this statistical model also amplifies the important role of F0 in distinguishing the aspirated stop from the lenis counterpart, at least in the case of labial /p^h/.

The mixed-effects logistic regression model without an interaction term (11) was also applied to the responses to the synthesized /k/, in order to examine the independent effects of each predictor variable. The output of the logistic model is shown in Table 4.12 and the associated figure.

Table 4.12. The output of the mixed-effects logistic regression model for the synthesized /k/.

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	1.021	1.01	
Number of obs: 288, groups: childid, 48				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-0.988934	1.211644	-0.816	0.414
Age	-0.027277	0.031467	-0.867	0.385
F0_difference	0.02423	0.006051	4.004	6.22e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

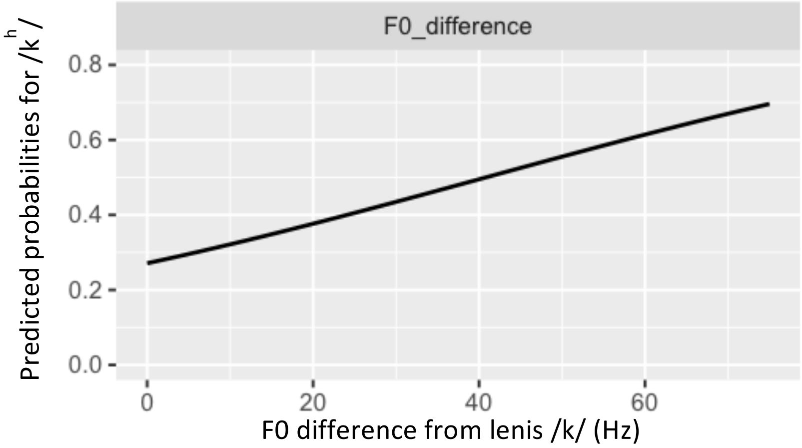

Figure 4.21. Relationship between F0 differences and the perception of aspirated /k^h/ with a logistic curve.

The multi-level logistic regression analysis also confirms that a relative F0 difference significantly affects the probability that the child participants perceive the stimuli as aspirated /k^h/ ($p < 0.001$). The perceptual patterns seemed irregular compared to the synthesized /p/, but the F0 difference is still positively related to the probabilities for perceiving /k^h/. As expected, there is no significant relationship between toddlers' age and their responses.

To find the interaction effect of age and F0 difference on the responses of /k^h/, the equation in (12) was applied to the synthesized /k/. The results of the regression with an interaction term are shown in the table below.

Table 4.13. The output of the mixed effects logistic regression model interaction effect of *Age* and *F0_ difference* in the case of the synthesized /k/.

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	1.054	1.027	
Number of obs: 288, groups: childid, 48				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	1.3661415	1.7607030	0.776	0.4378
Age	-0.0923640	0.0481093	-1.92	0.0549 .
F0_difference	-0.0316639	0.0310386	-1.02	0.3077
Age:F0_difference	0.0015297	0.0008438	1.813	0.0699 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

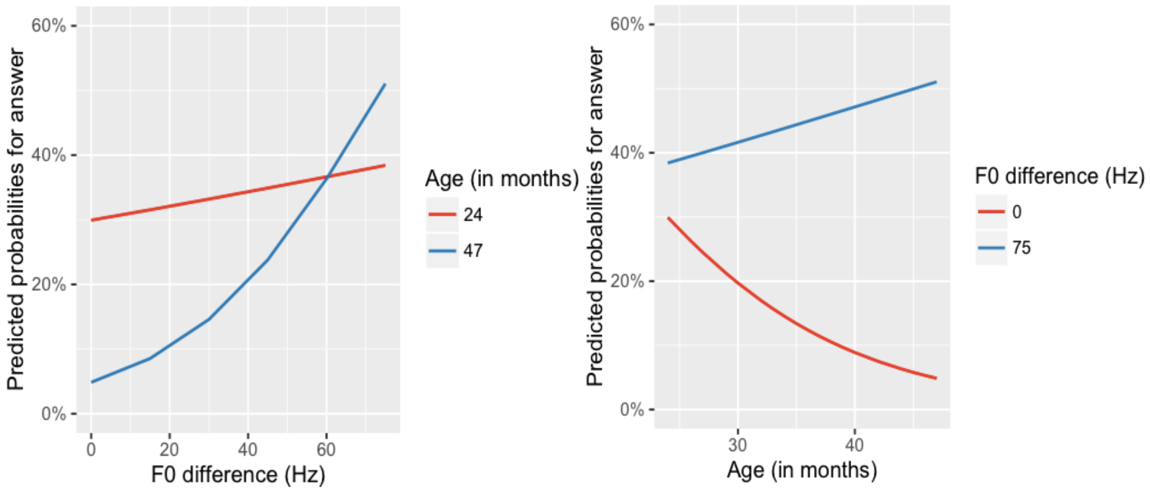

Figure 4.22. Interaction effect of *Age* and *F0_difference* on *Answer* of /k^h/.

The diverse irregular perceptual patterns for the synthesized /k/ are statistically confirmed. The fixed-effects coefficients in Table 4.13 show no significant effect of any predictors on the toddlers' responses (all *p*-values > 0.05). One interpretation is that the interaction of *Age* and *F0 difference* does not significantly affect the predicted probability for responses of /k^h/. That is, the probability of /k^h/ with each stimulus is not significantly related to children's age. Comparing to the result in Table 4.12, which showed that *F0_ difference* itself could be a significant predictor for the answer of /k^h/ without the age effect, it is notable that the higher rate of inconsistent responses was reflected in the analysis, so no significant effect of the interaction between *Age* and *F0 difference* could be yielded. Recalling the results in the case of synthesized /p/, there was a significant relationship of age to the toddlers' responses to different F0 differences from lenis /p/. Meanwhile, the multi-level regression analysis with the synthesized /k/ highlights that the same toddlers inconsistently and unstably identified the synthesized /k/ stimuli, which is represented in divergent, irregular perceptual patterns.

Next, the effect of age or F0 differences on the responses was revealed by conducting the same mixed-effects model with all of the obtained responses to both /p/ and /k/ stimuli. The regression result is presented in Table 4.14.

Table 4.14. The output of the mixed-effects logistic regression model for the effect of predictors on all responses to synthesized /p/ and /k/.

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	0.368	0.6067	
Number of obs: 576, groups: childid, 48				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-1.605252	0.780292	-2.057	0.0397 *
Age	-0.02051	0.020127	-1.019	0.3082
F0_difference	0.037517	0.004404	8.519	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

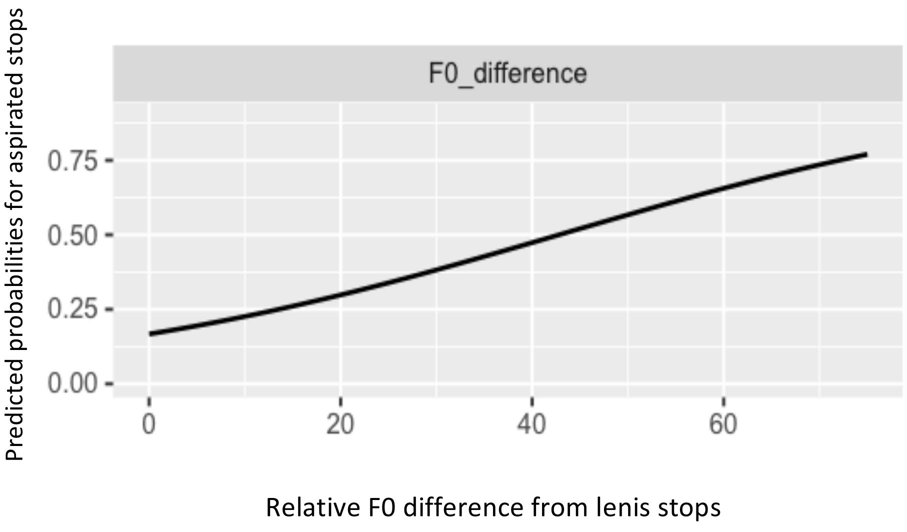

Figure 4.23. Relationship between F0 differences from lenis stops and the perception of aspirated stops with a logistic curve.

There is a significant effect of F0 differences on the probability of being perceived as an aspirated counterpart for stimuli ($p < 0.001$). This is a corresponding result to the previously performed logistic regression models, which revealed that the F0 difference each stimulus has significantly affects the toddlers' responses with both /p/ and /k/. As the F0 difference increases, the odds of the stimuli being identified as aspirated stops increases by 3.8% (*coef.* = 0.038). There is no age effect on the identification of aspirated stops, as predicted ($p > 0.05$).

The regression model from the equation in (13) was built in order to show how each POA pair is related to F0 difference and how significantly their interaction affects the toddlers' perceptual patterns.

$$(13)\ log\left(\frac{(Answer_{ij})}{1-(Answer_{ij})}\right) = \beta_0 + \beta_1 \times F0\_difference_{ij} + \beta_2 \times Velar_{ij}$$

$$+ \beta_3 \times F0\_difference_{ij} \times Velar_{ij} + \varepsilon_{ij} + \eta_i$$

(*Answer* = 1 if a child's response was aspirated stops, otherwise *Answer* =0)

Table 4.15. The output of the mixed-effects logistic regression model with an interaction term in (13).

Random effects:				
Groups	Name	Variance	Std. Dev.	
childid	(Intercept)	0.4448	0.6669	
Number of obs: 576, groups: childid, 48				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-3.173451	0.388430	-8.17	3.09e-16 ***
F0_difference	0.057653	0.007302	7.895	2.89e-15 ***
Velar	1.360725	0.455119	2.99	0.00277 **
F0_difference:Velar	-0.035818	0.009051	-3.954	7.61e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

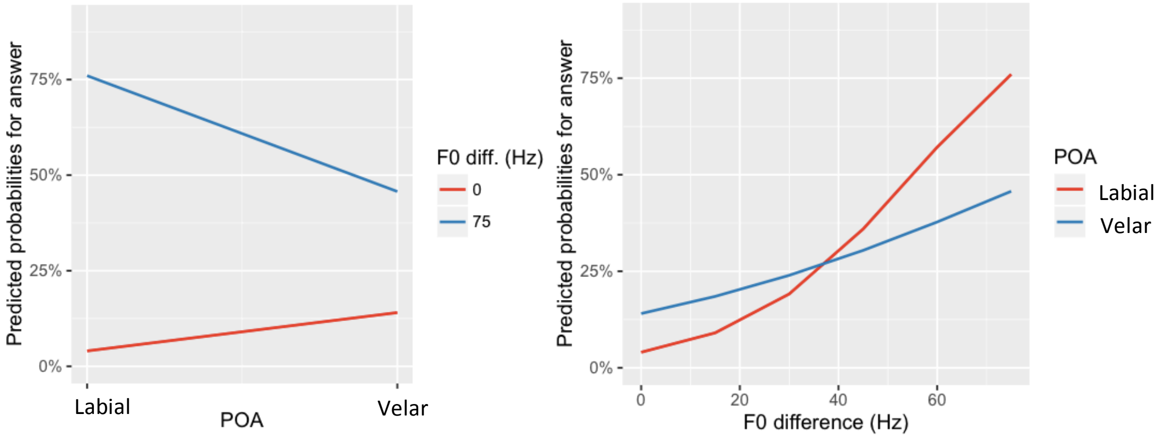

Figure 4.24. Interaction effect of *POA* and *F0 difference* on *Answer* of aspirated stop.

According to the output of the model in (13), the toddlers' responses are significantly affected by an F0 difference in both cases of labial and velar stop contrasts. More specifically, in the case of the synthesized /k/ stimuli, by F0 difference there is a 2.2% (=$e^{0.022}$) rate of increase in the odds ratio of aspirated stop /k^h/ as a response (*Answer* = 1), while a 5.9% (=$e^{0.058}$) rate of increase is predicted for a response of /p^h/. The increase rate of the odds ratio seems to represent a subtle change in the responses, even though the logistic regression model verified the significance of the F0 differences in different POA stop stimuli. The associated Figure 4.24 highlights the interpretation that two different POAs are significantly related to F0 differences, and that their interaction significantly influences the toddlers' responses in spite of the different size of effect depending on POA.

Chapter 5

Interrelation between child speech perception and production

In language development, the roles of perception and production are crucial. In spite of the importance of the interaction between speech production and perception, interrelated studies are limited, but it is essential to investigate the dynamics of speech production and perception especially for child language development. This chapter assesses the interactive development of an acoustic parameter, F0, in the phonemic speech processing of toddler speakers. The role of F0 in the Korean stop system is to phonetically distinguish aspirated and lenis stops in production and perception. With the recent trend of tonogenesis, the role of F0 has become essential, especially in discrimination between laryngeal contrasts in adult speech (Kang, 2014; Silva, 2006; Wright, 2007), but developmental patterns in terms of F0 have not been fully dealt with in linguistics studies (e.g., Kim, 2008; Kong *et al.*, 2011). Therefore, I hypothesize (1) that the role of F0 is determinant for the phonetic differentiation of lenis stops from aspirated stops, (2) that perceptually driven phonemic categorization for those stop categories occurs in the F0 dimension during the period of 2 to 4 years of age, and (3) that this phonemic categorization with phonetic details facilitates further development of articulatory distinction skills during that period in learning native contrasts.

With these hypotheses, this section is devoted to dealing with the phonetic gap between young children's production and perceptual patterns for the non-fortis stop contrast between lenis and aspirated stops. In addition, in relation to children's linguistic developmental stages, it is revealed how the observed phonetic differences in their acoustic space can decrease to some extent and that young children's articulatory distinctions truly correspond to the phonemic categories in their perceptual map.

Therefore, in this section, the results from the previous production experiment and the perception experiment with synthesized stimuli will be compared. In addition, the adult speakers' responses to the same synthesized sound tokens are the phonetic criterion to judge the children's acoustic development regarding F0 in stop contrasts. For a statistical analysis, a mixed-effects logistic regression model was conducted in order to determine the significant predictors for the phonetic differentiation of lenis and aspirated stops. Through this comparative analysis, the current study aims to provide experimental evidence for acoustic development in phonemic speech processing.

5.1. Labial lenis-aspirated stops

5.1.1. Overview

Through the perception experiment with synthesized /p/ stimuli, it was revealed that child speakers are quite sensitive to relative F0 differences, and that this acoustic difference leads to perceptual categorization of different phonemes in their acoustic space. Their perceptual thresholds for the phonemic categorization of aspirated /p^h/ vary relative to their different F0 development stages, which are accordingly affected by their age. In this circumstance, I needed criteria for which the stimulus truly sounds like an aspirated /p^h/, and for which another stimulus never sounds aspirated. Therefore, I decided to test the adult speakers who had already participated in the production experiment, with the same synthesized stimuli in the same condition as the toddlers. There were 14 female and 9 male adult speakers who were in their 30s and born in the 1980's. The figure below shows how each stimulus was perceptually identified by the child and adult participants. It is quite surprising that the perceptual identification of the adult speakers did not present a unified pattern except in one case: the stimulus with a 75-Hz difference from lenis /p/, 'p275', was perceived as aspirated /p^h/ without exception. Compared to the perception patterns of the adult speakers, the children's identification patterns were much more varied. In over 20% of the responses they reported that the 15-, 30-, and 45-Hz differences sounded like aspirated /p^h/, and in 50% of the responses to 'p260', they reported it as aspirated.

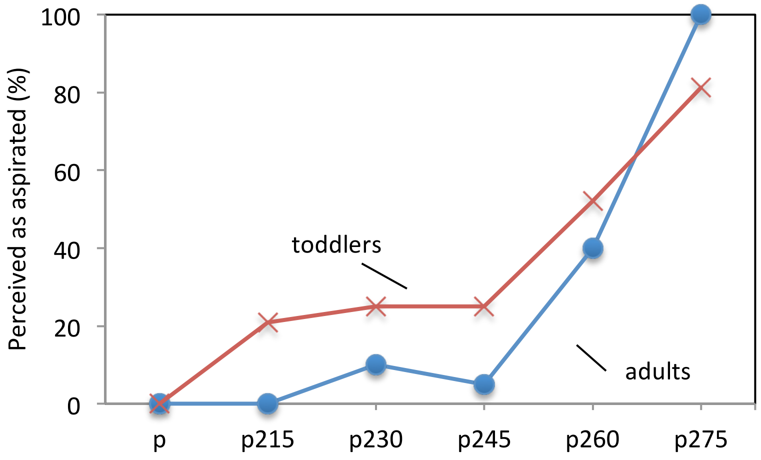

Figure 5.1. Results of the identification of synthesized /p/ by adult and toddler speakers. The numbers following ‘p’ denote the manipulated F0 values at vowel onset.

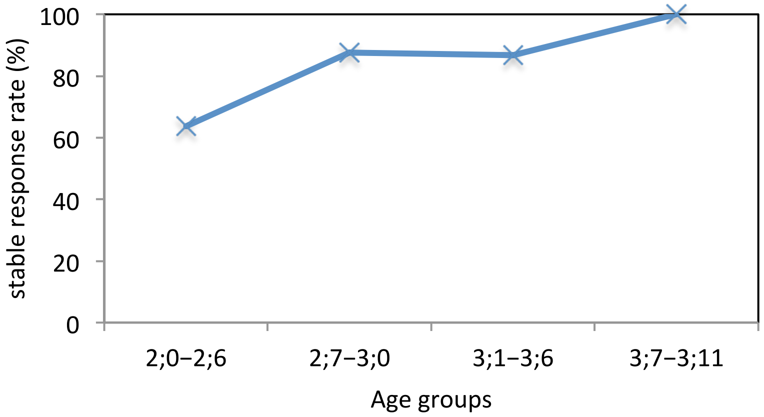

Figure 5.2. The ratio of stable responses to all obtained responses to synthesized /p/.

As can be seen in Figure 5.2, which shows the response stability of the children, only around 60% of the children in the youngest age group had a robust phonetic criterion to determine aspirated stops. It is an indication that during the period of age 2;0 to 2;6, toddlers’ perception is easily affected by their lexicon and by the given context, rather than being dependent on fine phonetic details. The fact that response stability increases as toddlers grow would indicate that their perceptual sensitivity to F0 decreases, so their perception of the lenis-aspirated stop distinction becomes consistent and stabilized to some extent.

To compare the distance in acoustic space between /p/ and /p^h/ in children's productions, the estimated phonemic categories are presented with standard errors for the four different age groups in Figure 5.3. The youngest age group's productions show the largest standard errors both in VOT and F0, which means their articulatory distinction for the stop contrast has considerable variation in its phonetic values. In particular, the dimension of F0 has the highest standard errors, which indicates an F0 merger in which there is a good deal of overlap in the F0 values of lenis and aspirated stops. Before the age of 3 years, the articulatory distinction between lenis and aspirated stops largely depends on relative VOT differences rather than a difference in F0.

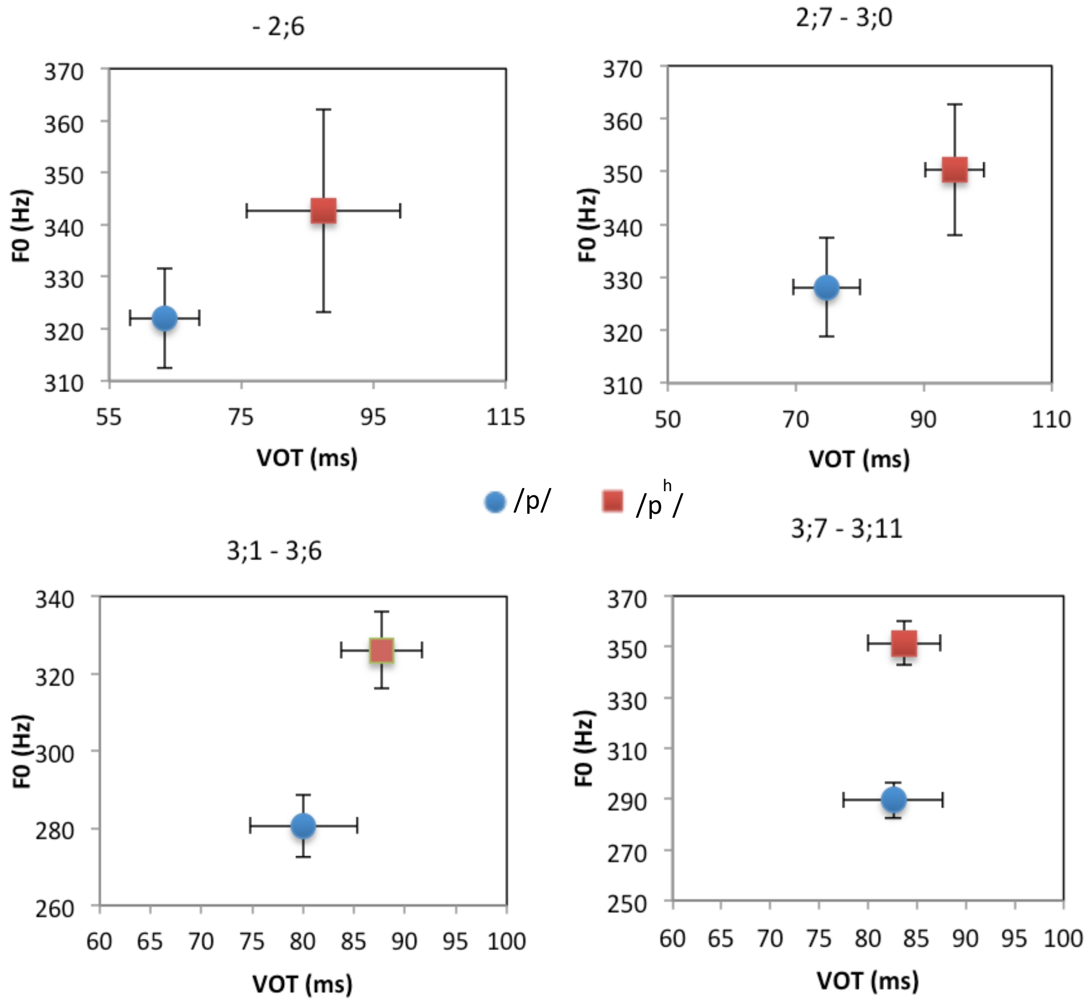

Figure 5.3. Phonemic categories of labial lenis and aspirated stops in the dimensions of VOT (ms) and F0 (Hz) by age group. Error bars represent standard errors.

After the age of 3 years, the VOT merger is clearly noticeable, while it is simultaneously observed that there is a relatively long acoustic distance in the dimension of F0 between /p/ and /p^h/. The oldest group showed a complete VOT merger, but the two stop categories could be differentiated by F0 values. Over time, the two stop categories are clearly formed into two different ranges of F0. It should be speculated why the older children did not use VOT differentiation even though they already acquired the way to produce substantial VOT differences between lenis and aspirated stops. It is interesting that the children's F0 developmental patterns recapitulate the recent linguistic process of tonogenesis in Seoul Korean; VOT merger is observed while F0 distinction is enhanced between lenis and aspirated stops (Kang, 2014; Silva, 2006; Wright, 2007).

5.1.2. Statistical analysis

Using a mixed-effects linear regression model (Gelman & Hill, 2006) in R by implementing the *lme4* package (Bates *et al.*, 2015), it is revealed how toddler speakers differentiated VOT or F0 for producing the lenis stop /p/. In this model, I only dealt with the phonetic distinction between the non-fortis stops /p/ and /p^{h}/ in relation to toddlers' age.

The equation used is as follows. Lenis stop /p/ was the reference category.

(14) $VOT_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Asp_{ij} + \beta_3 \times Age_i \times Asp_{ij} + \varepsilon_{ij} + \eta_i$

Table 5.1. The output of the mixed-effects linear regression model based on (14).

Random effects:				
Groups	Name	Variance	Std.Dev.	
child_id	(Intercept)	481.5	21.94	
Residual		1178.7	34.33	
Number of obs: 534, groups: child_id, 58				
Fixed effects:				
	Estimate	Std.Error	t value	Pr (>|t|)
(Intercept)	52.14275	18.413	2.832	0.00606 **
Age	0.64739	0.49107	1.318	0.19185
Aspirated	12.87169	16.56375	0.777	0.43751
Age:Aspirated	-0.07068	0.433	-0.163	0.87042

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The equation in (14) was designed in order to determine how much stop category or toddlers' age would affect VOT changes. With this multi-level regression model, it can be seen that there is not a significant effect of *Age* on VOT changes in each non-fortis stop category (*p*-values > 0.05). Recalling the production patterns of the toddlers, the younger ones were likely to effectively use VOT as an acoustic tool for distinction between /p/ and /p^h/, while the relatively older ones used F0 to differentiate the two stop phonemes. This contrastive use of the acoustic parameters by the toddlers brings out the insignificant effect of the relationship between age and stop category on VOT changes.

An identical mixed-effects model was applied to F0 changes. For this modeling, the equation in (15) was constructed. Essentially, this model was used to determine the statistical significance of the predictors of toddler speakers' age and stop categories on F0 variations in toddlers' production of /p/ and /p^h/. The lenis stop /p/ was the reference category.

(15) $F0_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Asp_{ij} + \beta_3 \times Age_i \times Asp_{ij} + \varepsilon_{ij} + \eta_i$

Table 5.2. The output of the mixed-effects linear regression model based on (15).

Random effects:				
Groups	Name	Variance	Std.Dev.	
child_id	(Intercept)	1191	34.52	
Residual		5115	71.52	
Number of obs: 534, groups: child_id, 58				
Fixed effects:				
	Estimate	Std.Error	t value	Pr (>|t|)
(Intercept)	360.308	32.686	11.023	<2e-16 ***
Age	-1.6468	0.8698	-1.893	0.0617 .
Aspirated	-29.8155	34.2811	-0.87	0.3849
Age:Aspirated	1.8468	0.8966	2.06	0.0400 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

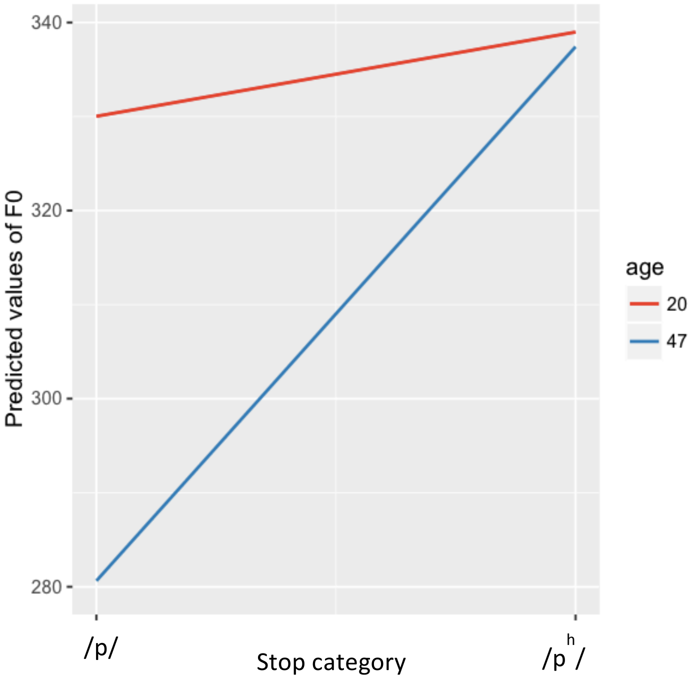

Figure 5.4. Interaction effect of age (in months) and phonation type on F0 from the mixed-effects linear model in (15).

Unlike the model for VOT in (14), it is observed that the interaction between children’s age and aspirated /p^{h}/ is a significantly effective predictor for F0 changes ($p < 0.05$), while the interaction between *Age* and lenis /p/ is relatively weak ($p < 0.1$). By age difference, the predicted F0 change for /p^h/ compared to F0 in /p/ would be around 0.2 Hz (=−1.6468 + 1.8468). This would not be considered a large phonetic change, but it is important that there are significant effects of stop manner on F0 with age. With age, F0 changes are observed to become preferable to VOT changes when children phonetically differentiate aspirated stops from lenis stops. Figure 5.4 presents the F0 differences in production of the youngest and oldest child, which shows that the oldest child produces substantial F0 difference between /p/ and /p^h/. This production pattern is compatible with the finding from previous studies that VOT is acquired earlier, but that once F0 is acquired, it is used preferentially, especially for the differentiation of lenis stops (Kim, 2008; Kim & Stoel-Gammon, 2009; Kong, 2009; Kong *et al.*, 2011).

A mixed-effects logistic regression model should be useful to assess the effects of VOT or F0 on the production of the aspirated stop. This multi-level logistic model was conducted using the *glmer* function in the *lme4* package (Bates *et al.*, 2015) in R. The equation in (16) was designed for this logistic model and was applied identically to the adult speakers' data and the toddlers' data. The mixed-effects logistic model used *VOT* and *F0* as the fixed effects predictors, and the by-speaker adjustment to intercept was a random effect. Through this statistical method, it can be shown how similarly or differently the acoustic parameters operate to produce the phonetic distinction of aspirated stop /p^h/ in the adult and toddler speakers' production.

(16) $log\left(\frac{/Aspiration_{ij}/}{1-/Aspiration_{ij}/}\right) = \beta_0 + \beta_1 \times VOT_{ij} + \beta_2 \times F0_{ij} + \varepsilon_{ij} + \eta_i$

(*Aspiration* = 1 if aspirated stop /p^h/, otherwise *Aspiration* =0)

Table 5.3. The output of the mixed-effects logistic regression model for male adult speakers' production.

Random effects:				
Groups	Name	Variance	Std.Dev.	
adult_id	(Intercept)	10.61	3.257	
Number of obs: 108, groups: adult_id, 9				
Fixed effects:				
	Estimate	Std.Error	z value	Pr (>|t|)
(Intercept)	-26.01886	7.08883	-3.67	0.000242 ***
VOT	0.13751	0.00752	1.774	0.76105 .
F0	0.11234	0.2914	3.855	0.000116 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1				

Table 5.4. The output of the mixed-effects logistic regression model for female adult speakers' production.

Random effects:				
Groups	Name	Variance	Std.Dev.	
adult_id	(Intercept)	22.91	4.786	
Number of obs: 164, groups: adult_id, 14				
Fixed effects:				
	Estimate	Std.Error	z value	Pr (>|t|)
(Intercept)	-48.03688	10.1804	-4.719	2.38e-06 ***
VOT	0.09614	0.03506	2.743	0.0061 **
F0	0.17742	0.03875	4.578	4.69e-06 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1				

Due to the relative difference in F0 values between female and male adult speakers, the two gender groups were analyzed separately. According to the results, the male adult speakers' productions of /p/ and /p^h/ are significantly differentiated by F0 ($p < 0.001$). On the other hand, the female speakers tended to differentiate /p/ and /p^h/ with a combination of F0 and VOT (p-values < 0.01). However, the impact of the contribution of VOT seems much smaller than that of F0 since the fixed effects coefficients for VOT and F0 are reported as 0.096 and 0.177 respectively. The predicted probabilities for aspirated /p^h/ in relation to F0 and VOT are as follows.

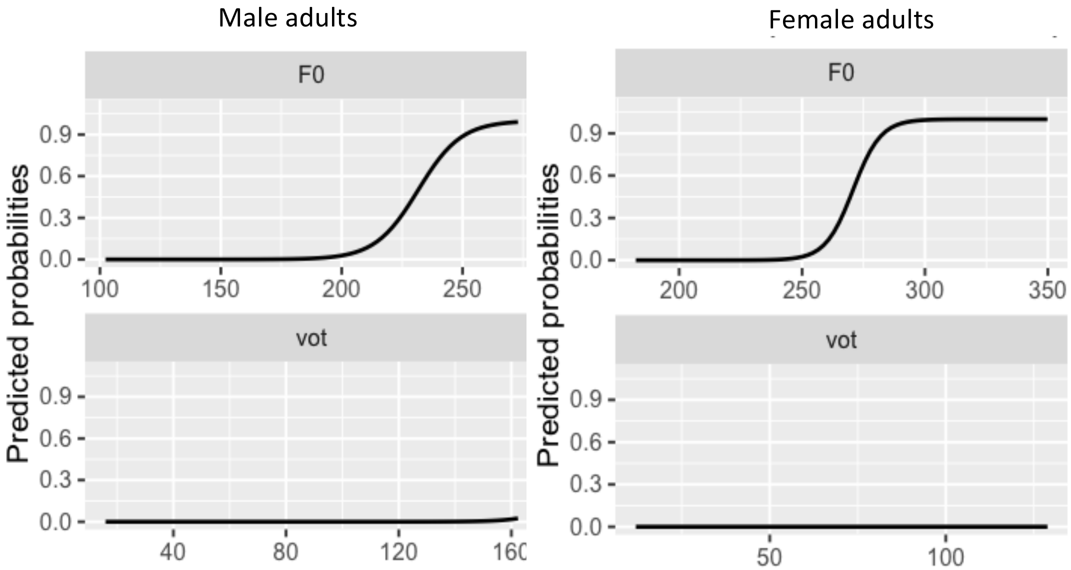

Figure 5.5. Predicted probabilities for /p^h/ of male (left panel) and female (right panel) adult speakers, calculated by the mixed effects logistic regression model in (16).

As the figure illustrates, in both cases, the role of VOT does not seem very effective while F0 is significantly related to the production of aspirated /p^h/. This phenomenon is consistent with the tonogenetic sound change as F0 differentiation becomes critical for the distinction between lenis and aspirated stops.

The same equation was applied to the toddler speakers' data. The productions of the youngest age group had no significant phonetic differentiation for /p/ and /p^h/. Table 5.5 suggests that VOT and F0 do not affect the production of /p^h/, which means that the toddlers in the youngest age group did not show any substantial phonetic differences in the production of the two stop categories (*p*-values > 0.1). The logistic regression model reported that the estimated random effect variance is 0, which can be interpreted as no effect on the probability for aspirated /pʰ/ due to individual differences, since the number of subjects in this data matrix is considered too small to analyze (*n* = 14).

Table 5.5. The output of the mixed-effects logistic regression model for the age group (1;8-2;6).

Random effects:				
Groups	Name	Variance	Std. Dev.	
child_id	(Intercept)	0	0	
Number of obs: 99, groups: child_id, 14				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-1.297806	0.893	-1.453	0.146
VOT	0.00508	0.004755	1.069	0.285
F0	0.00207	0.002625	0.82	0.421

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Table 5.6. The output of the mixed-effects logistic regression model for the age group (2;7–3;0).

Random effects:				
Groups	Name	Variance	Std. Dev.	
child_id	(Intercept)	0	0	
Number of obs: 136, groups: child_id, 13				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-2.45111	0.997248	-2.458	0.01398 *
VOT	0.015461	0.005767	2.681	0.00734 **
F0	0.003439	0.002452	1.403	0.16074

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The second youngest age group’s (2;7-3;0) results reveal a somewhat different production pattern in Table 5.6. There is no significant effect of F0 (*p* > 0.05) on the probabilities for /pʰ/, but VOT begins to operate as an important acoustic tool to distinguish the two stop phonemes in their production ($p < 0.01$). It is interesting that the toddlers over 2;7 seem to start to differentiate the two contrastive phonemes in articulation using phonetic details. It is expected that for 3-year-old children the acoustic space with respect to F0 has not yet developed enough to differentiate the lenis stop, so it seems that they tried to distinguish them alternatively with VOT. This production pattern represents a commonly observed phenomenon, in which lenis stops are replaced with fortis counterparts in child speech.

The third age group (3;1–3;6) produced /p^h/ using F0 differentiation. F0 is a significantly effective predictor to produce /p^h/ in their speech ($p < 0.01$), while the role of VOT seems to diminish ($p > 0.05$) compared to the second youngest age group's results. Recalling that the second age group showed substantial VOT differentiation for the lenis-aspirated distinction, it is a surprising linguistic development that F0 differentiation appears in the production system of the third age group. This also indicates that the perceptual system has developed in the dimension of F0 between these ages, which should promote more accurate perception with F0. Therefore, the period of 3;1 to 3;6 seems to be a time for critical F0 development that affects the production system.

Table 5.7. The output of the mixed-effects logistic regression model for the age group (3;1–3;6).

Random effects:				
Groups	Name	Variance	Std. Dev.	
child_id	(Intercept)	0	0	
Number of obs: 158, groups: child_id, 17				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-2.229105	0.773123	-2.883	0.00394 **
VOT	0.004949	0.00444	1.115	0.26507
F0	0.006472	0.00222	2.915	0.00355 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The output of a logistic regression model for the oldest age group suggests that the role of F0 becomes more crucial in distinguishing aspirated /p^h/ from /p/. The fixed effects coefficient for F0 is still much smaller than for female adult speakers' production (*coef.* = 0.016 vs. 0.177 respectively), but the size of the effect is the biggest compared to the results from the three younger age groups ($p < 0.001$). For the children who are over 3 years old, in distinguishing /p^h/ from /p/, it seems that the role of VOT is minimized while F0 functions critically, similar to the pattern of the adult speakers. This resembles the male adult speakers' distinction pattern, even though the degree of the effect of F0 still does not reach that of adult speech. This is an indication that F0 development has occurred in the children's acoustic space and that phonemic categorization has also occurred in the dimension of F0, so that they can recognize and distinguish the phonological contrasts in the Korean stop system.

Table 5.8. The output of the mixed-effects logistic regression model for the age group (3;7 – 3;11).

Random effects:				
Groups	Name	Variance	Std. Dev.	
child_id	(Intercept)	0.2765	0.5258	
Number of obs: 141, groups: child_id, 14				
Fixed effects:				
	Estimate	Std. Error	z value	Pr (>|t|)
(Intercept)	-5.496047	1.416501	-3.88	0.000104 ***
VOT	0.001869	0.005248	0.356	0.721769
F0	0.016828	0.004143	4.061	4.88e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

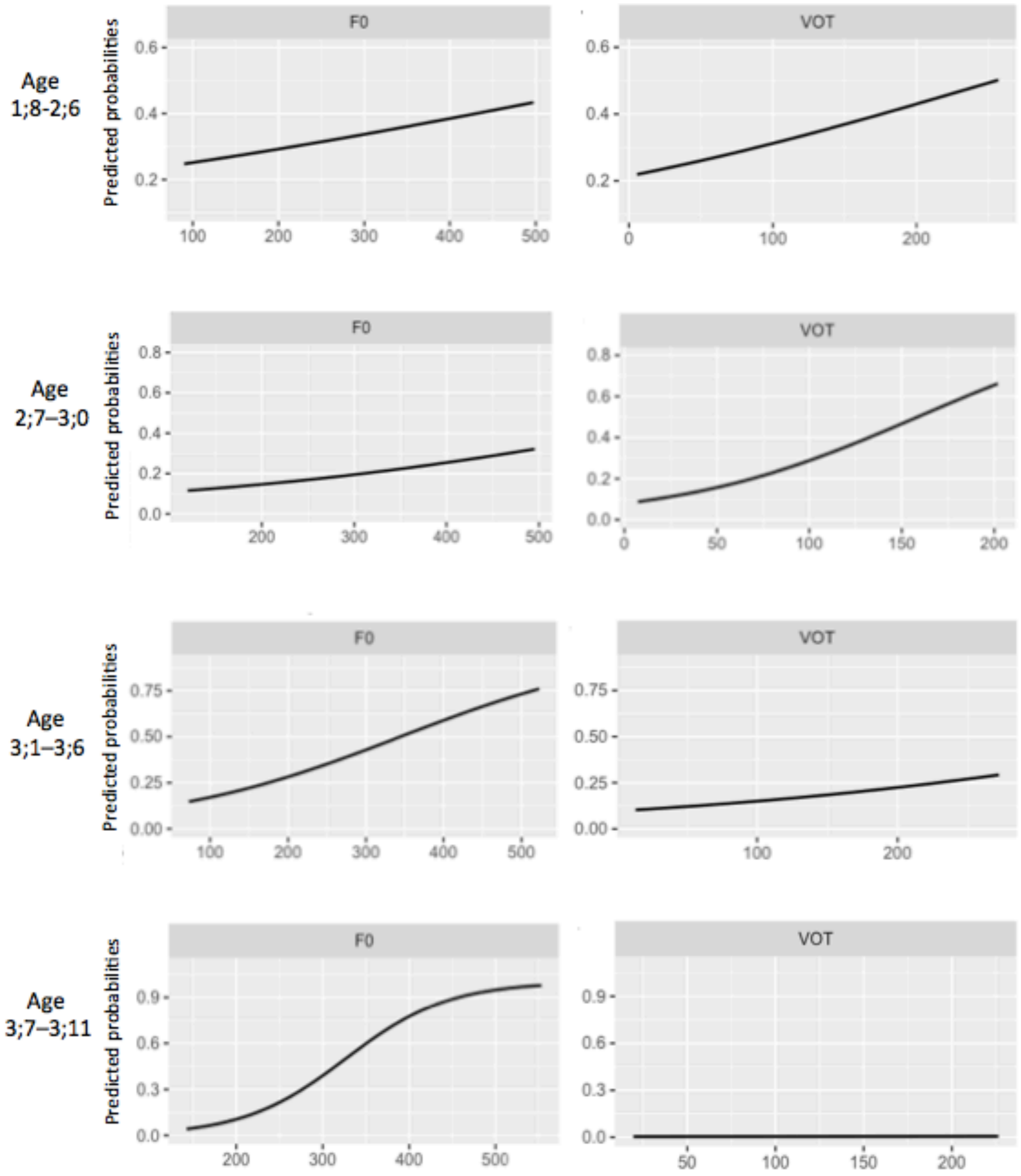

Figure 5.6. Probabilities for /p^h/ by F0 and VOT difference in productions by four child age groups. (Top row: 1;8-2;6, second row: 2;7-3;0, third row: 3;1-3;6, bottom row: 3;7-3;11).

The logistic curves in relation to the fixed effects predictors, F0 and VOT, for producing /p^h/ with four age groups are illustrated in Figure 5.6. Until the age of 3 years, VOT differences dominate in producing the acoustic characteristics of lenis and aspirated stops. However, F0 overlaps are easily found in the production of lenis and aspirated stops. This indicates a possibility that younger children tend to replace /p/ with fortis /p'/. After the age of 3 years, F0 begins to contribute substantial phonetic differences between aspirated and lenis stops. Consistently, it is observed that the relation between VOT and the probabilities for /p^h/ becomes very diminished with age as the slope of the logistic curve becomes flat. Conversely, F0 changes largely affect the probabilities for /p^h/ with age, as the slope of the logistic curve becomes steeper, which means the dependence of the production of /p^h/ on F0 increases.

5.2. Velar lenis-aspirated stops

5.2.1. Overview

The previous perception experiment with the synthesized /k/ stimuli showed irregular perceptual identification patterns by the toddler speakers. Figure 5.7 compares how the toddler and adult speakers perceptually identified each synthesized /k/ stimulus.

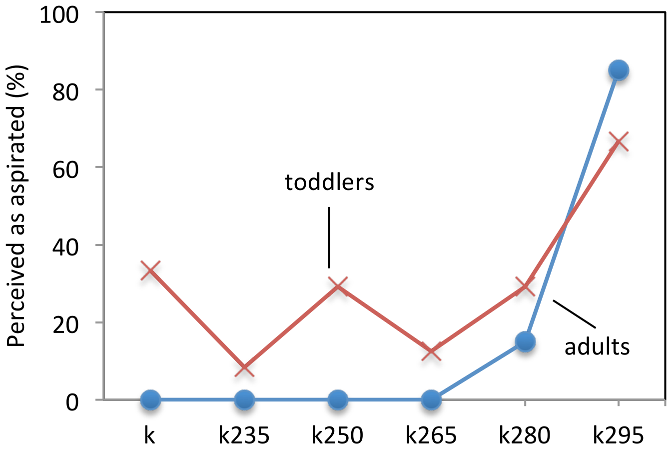

Figure 5.7. Results of the identification of synthesized /k/ by toddler and adult speakers. The numbers following ‘k’ denote the manipulated F0 values at vowel onset.

Recall that about 40% of the toddlers perceived plain /k/ with 220 Hz as aspirated counterpart /k^h/, while only 29.5% perceived /k/ with 280 Hz at vowel onset (which is a 60 Hz relative difference from plain /k/) as /k^h/. The perception pattern by the adult speakers seems much simpler compared to the toddlers’ responses; until they heard the stimulus with the relative difference of 60 Hz, they perceived lenis /k/ for smaller F0 differences. The only two stimuli with a 60 Hz or greater F0 difference, ‘k280’ and ‘k295’, were identified as aspirated /k^h/ by adults. Adult speakers have a certain phonetic standard to perceptually categorize aspirated stops compared to their lenis counterparts, and their response pattern suggests that the perceptual thresholds for /k^h/ need much higher F0 values (around 60 to 75 Hz) than lenis /k/.

Response stability gradually grows with age, as shown in Figure 5.8, even though it seems difficult to show a clear trajectory of the toddler speakers’ perceptual development with F0 regarding velar stop contrasts. This indicates that toddler speakers come to have their own phonetic standards for phonemic categorization as they grow older. Still, it is notable that the oldest age group shows less than 100% response stability, which implies that acquisition of the velar stop contrasts is delayed compared to the labial stop contrasts (recalling the 100% response stability of the oldest age group in the case of /p/ stimuli).

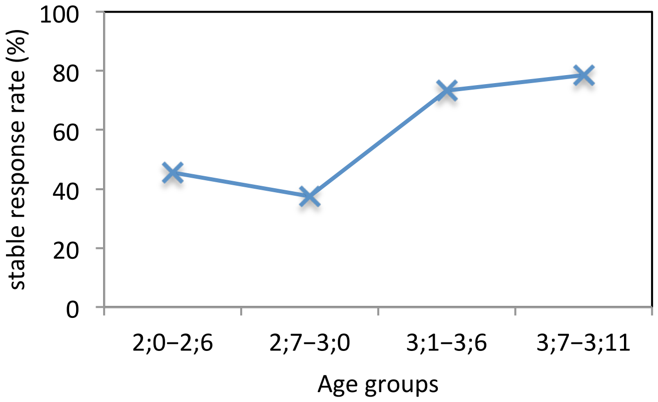

Figure 5.8. The ratio of consistent responses to all obtained responses to synthesized /k/ by age group.

One of the important assumptions in speech processing is that perceptual development occurs prior to the development of produciton. In order to compare the toddlers' acoustic space regarding phonemic categories for /k/ and /k^h/, the acoustic distance of the two stop categories was calculated from their production.

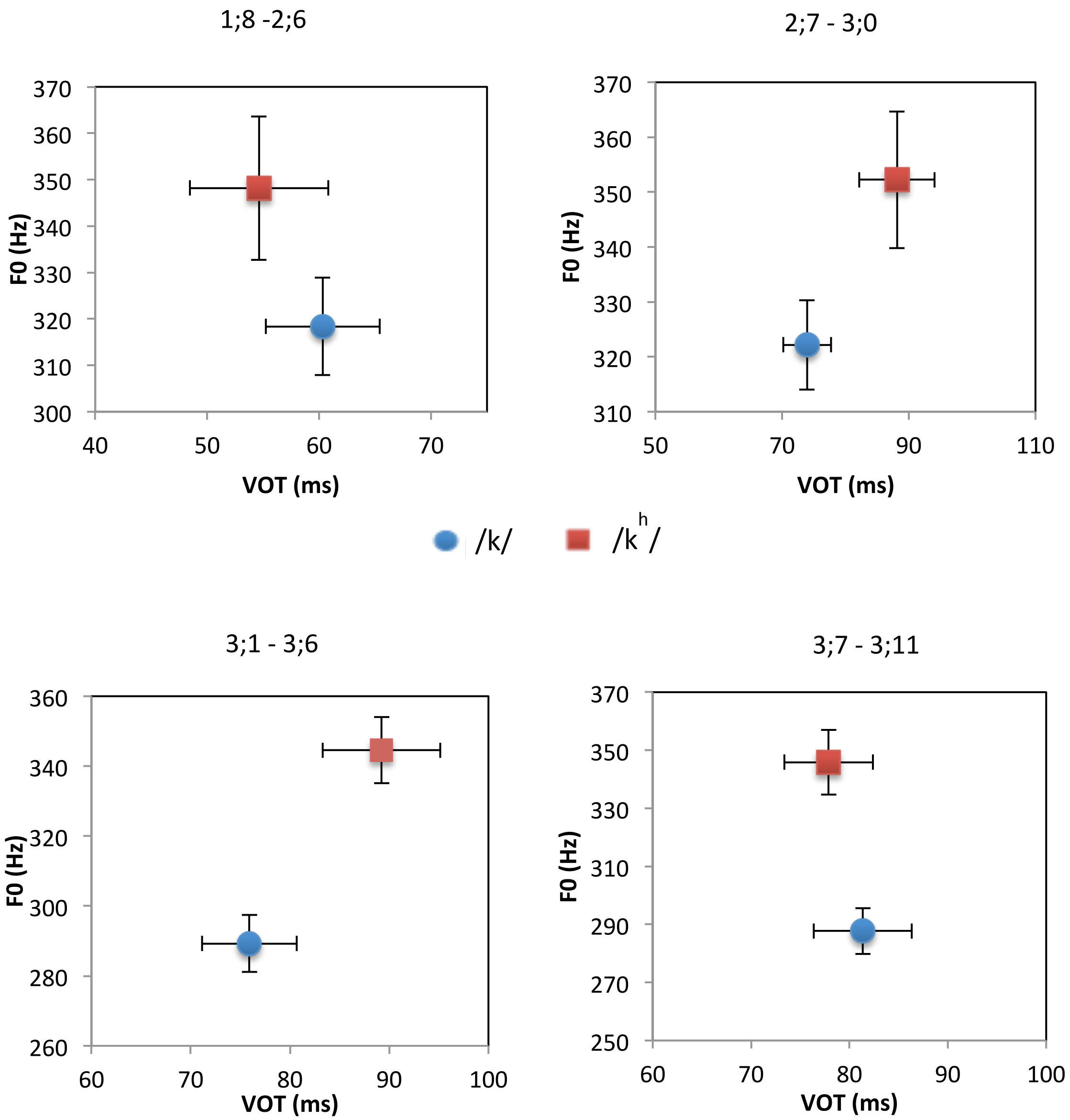

Figure 5.9. Phonemic categories of velar lenis and aspirated stops in the dimension of VOT (ms) and F0 (Hz) by age group. Error bars represent standard errors.

It is interesting that the young children's phonetic differentiation of /k/ and /kʰ/ varies depending on age and VOT and F0 operate complementarily to distinguish those stop categories. In the production of the youngest age group (2;0-2;6), F0 and VOT values for the two stops have large standard errors, which indicates that their production patterns were diverse and not standardized. Their VOT values for /k^h/ and /k/ are reversed since those for /k/ have larger values. Notice that the VOT values overlap in their standard errors, which means that there is a possibility that the youngest children have replaced both /k^h/ and /k/ with fortis /k'/. The second youngest age group's production shows that the former pattern is reversed, such that /k/ has shorter VOT values compared to /k^h/, while maintaining the same degree of F0 difference between the two stop categories. The third group shows decrease in overlap of VOT, but it is expected that the actual VOT values could vary. On the other hand, F0 difference increases with smaller standard errors. This means that during the period of 3;0 to 3;6, the use of F0 has been quite stabilized and young children have recognized F0 as a crucial mechanism for the distinction between /k^h/ and /k/.

The production of the oldest age group shows a complete VOT merger, while the F0 distinction was enhanced. The difference from the third age group is that the redundant VOT distinction diminished. This resembles a tonogenetic sound change process in which F0 differences are enhanced with a relative loss of VOT differences. This phonetic trade-off between F0 and VOT can be evidence for tonogenesis in progress in Seoul Korean, but it is still vague what causes this phenomenon in toddler speech, since their linguistic input should be adult productions, which have been undergoing tonogenesis so their productions possibly have F0 differentiation with/without VOT differentiation for lenis and aspirated stops. It should be more important to note that toddler speakers eventually become to use F0 dimensional differentiation for lenis and aspirated stop contrasts.

In the oldest age group's productions in the VOT dimension, the lenis stop has slightly shorter VOT values than the aspirated stop, which is not accurate. This phenomenon itself indicates that the phonetic differentiation of /k/ and /k^h/ would be difficult even for the oldest children and that their productions are confusing and not mature, which implies that their perception has not been stabilized yet. The relative F0 difference between the two phonemes is consistently observed, but the production patterns are not perfectly mature, so it is understandable that the stable perception of F0 differences would have been challenging for the child speakers.

5.2.2. Statistical analysis

This section aims to help understand the role of F0 development in the production of a velar stop contrast using a statistical model. The equation in (17) for a mixed-effects linear regression predicts how significantly VOT affects the phonation types in relation to children's age. The fixed effects predictors are *Aspirated* /k^h/ and *Age*, and by-speaker adjustment to intercept is a random effect. *Lenis* /k/ was the reference category.

(17) $VOT_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Asp_{ij} + \beta_3 \times Age_i \times Asp_{ij} + \varepsilon_{ij} + \eta_i$

Table 5.9. The output of the mixed-effects linear regression model based on (17).

Random effects:				
Groups	Name	Variance	Std.Dev.	
child_id	(Intercept)	358	18.92	
Residual		1281	35.79	
Number of obs: 498, groups child_id, 58				
Fixed effects:				
	Estimate	Std.Error	t value	Pr (>|t|)
(Intercept)	46.1843	19.025	2.427	0.0168 *
Age	0.7184	0.5085	1.413	0.1605
Aspirated	19.9069	18.8693	1.055	0.2921
Age:Aspirated	-0.3242	0.505	-0.642	0.5213

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

As expected, the VOT values are not significantly influenced by phonation type nor is a significant relationship between children’s age and phonation type. However, the F0 values are significantly dependent on phonation type and are also significantly related to children’s age (p-values < 0.05) as shown in Table 5.10.

(18) $F0_{ij} = \beta_0 + \beta_1 \times Age_i + \beta_2 \times Asp_{ij} + \beta_3 \times Age_i \times Asp_{ij} + \varepsilon_{ij} + \eta_i$

Table 5.10. The output of the mixed-effects linear regression model based on (18).

Random effects:				
Groups	Name	Variance	Std.Dev.	
child_id	(Intercept)	347.7	18.65	
Residual		5276.4	72.64	
Number of obs: 498, groups: child_id, 58				
Fixed effects:				
	Estimate	Std.Error	t value	Pr (>|t|)
(Intercept)	390.7623	30.7212	12.579	< 2e-16 ***
Age	-2.3714	0.8206	-2.734	0.00699 **
Aspirated	-41.1832	37.8281	-1.089	0.27695
Age:Aspirated	2.2919	1.0122	2.249	0.02509 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

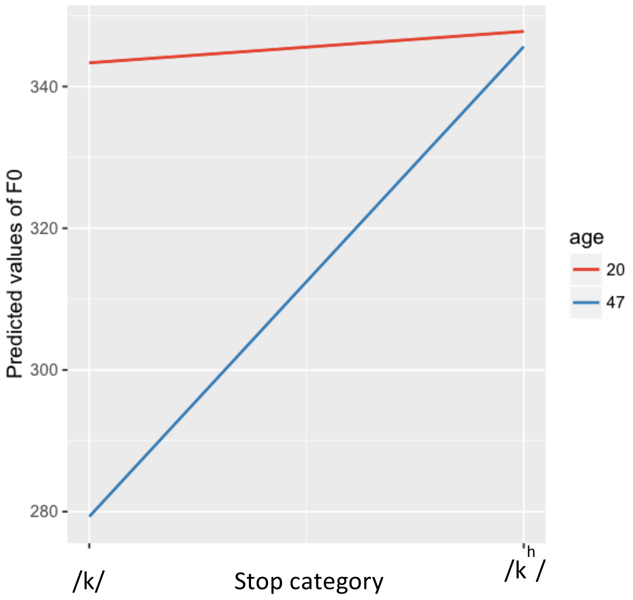

Figure 5.10. Interaction effect of age (in months) and phonation type on F0 from the mixed-effects linear model in (18).

The estimated F0 change for /k/ is around −2.37 Hz by age (in months) difference. The interaction effect of age and stop category on F0 change is illustrated in Figure 5.9. The youngest child's (20 months=1;8) F0 difference depending on stop phoneme is relatively subtle compared to the F0 difference between /k/ and /k^h/ by the oldest child (47 months=3;11). Utilizing F0 as an acoustic tool to distinguish between lenis and aspirated stops in production should be directly related to toddlers' perceptual system, in which they can recognize a certain range of F0 differences so the stop contrast can be appropriately and categorically perceived in their acoustic space.

The second statistical method used was a mixed-effects logistic regression. Modeling the logistic regression makes it possible to predict how VOT or F0 interacts with each phonation type produced by the four different age groups. The equation used is illustrated in (19).

(19) $$log\left(\frac{/Aspiration_{ij}/}{1-/Aspiration_{ij}/}\right) = \beta_0 + \beta_1 \times VOT_{ij} + \beta_2 \times F0_{ij} + \varepsilon_{ij} + \eta_i$$

(If a production was /k^h/, *Aspiration* = 1, otherwise *Aspiration* = 0)

Productions of adult speakers were analyzed separately by gender. The results for the male speakers are presented in Table 5.11. Similarly to the results from the labial stop contrast, the role of VOT is not significant enough to determine /k^h/ ($p > 0.1$), while F0 significantly affects the production of the aspirated stop ($p < 0.01$).

Table 5.11. The output of the mixed-effects logistic regression for the production of /k^h/ by male adult speakers.

Random effects:				
Groups	Name	Variance	Std.Dev.	
adult_id	(Intercept)	24.76	4.975	
Number of obs: 108, groups: adult_id, 9				
Fixed effects:				
	Estimate	Std.Error	z value	Pr (>|t|)
(Intercept)	-41.090526	13.311482	-3.087	0.00202 **
VOT	-0.001356	0.044612	-0.03	0.97575
F0	0.286766	0.095066	3.017	0.00256 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The results of the female speakers are the same. F0 is significantly related ($p < 0.001$) to phonation type rather than VOT ($p < 0.1$). In this instance, however, the impact of the effects of F0 are quite different; in the case of the male speakers, there is a 33% increase (=$e^{0.286}$) in the odds ratio of /k^h/ by F0 difference (Hz), whereas female speakers affect as much as 15% increase (=$e^{0.144}$) in the odds ratio of /k^h/. Recalling that in distinction between /p/-/p^h/, both VOT and F0 significantly affect the production of female adult speakers, it is notable that there is no significant VOT differentiation for the /k/-/k^h/ distinction.

Table 5.12. The output of the mixed-effects logistic regression for the production of /k^h/ by female adult speakers.

Random effects:				
Groups	Name	Variance	Std.Dev.	
adult_id	(Intercept)	5.83	2.415	
Number of obs: 166, groups: adult_id, 14				
Fixed effects:				
	Estimate	Std.Error	z value	Pr (>|t|)
(Intercept)	-37.16254	7.33771	-5.065	4.09e-07 ***
VOT	0.04703	0.02588	1.817	0.0692 .
F0	0.14479	0.02937	4.93	8.21e-07 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1				

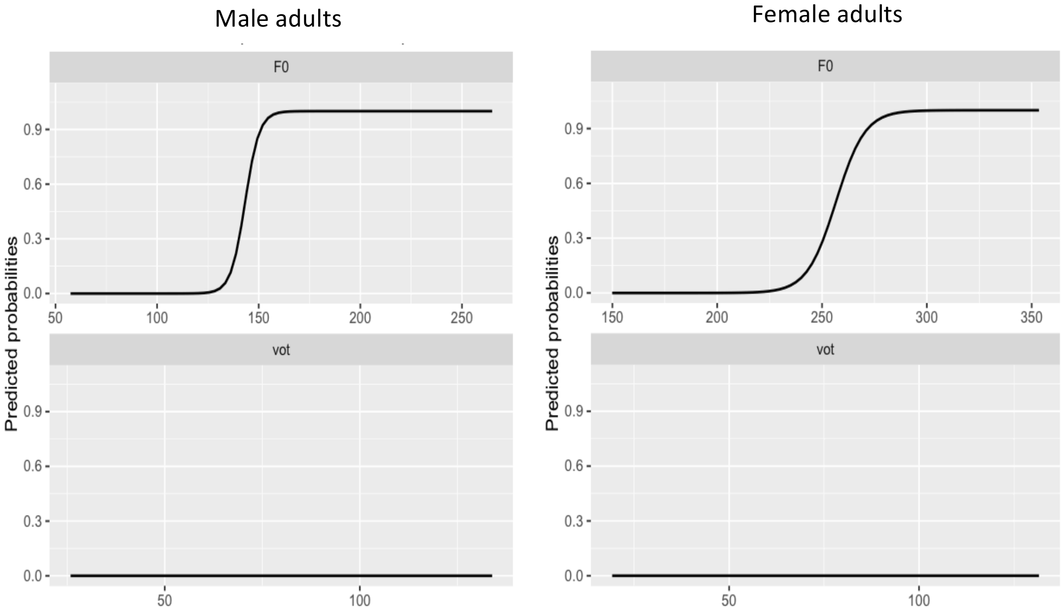

Figure 5.11. Predicted probabilities for /k^h/ of male (left panel) and female (right panel) adult speakers, which are estimated through the mixed-effects logistic regression model in (19).

The logistic curves from the multi-level regression model are shown in Figure 5.11. As expected, VOT values barely affect the probabilities for /k^{h}/, while the impact of F0 is much bigger and directly related in both of male and female adult productions.

The children's productions were divided into four different age groups, and each group was analyzed separately for effective comparisons. The youngest age group's results show that their articulatory confusion between /k/ and /k^{h}/ seems related to both VOT and F0, since neither of them has a significant effect on phonetic differentiation. This indicates that the two stop categories were not phonetically distinguishable in the youngest group's actual productions.

Table 5.13. The output of the mixed-effects logistic regression for the youngest age group (1;8-2;6).

Random effects:				
Groups	Name	Variance	Std.Dev.	
child_id	(Intercept)	1.55e-16	1.6507	
Number of obs: 97, groups: child_id, 14				
Fixed effects:				
	Estimate	Std.Error	z value	Pr (>|t|)
(Intercept)	-0.783768	0.921596	-0.85	0.395
VOT	-0.002414	0.005582	-0.432	0.665
F0	0.003307	0.002559	1.292	0.196
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1				

The second age group (2;7–3;0) produced significant phonetic differentiation for the distinction between /k^{h}/ and /k/. Both VOT and F0 significantly influenced the phonation types (*p*-values < 0.05), which indicates that there are significantly observable VOT and F0 differences between their productions of /k/ and /k^{h}/. According to the results, before 3 years of age, the children tend to differentiate F0 and VOT values to discriminate the velar contrast. This is a big difference from the results of the labial stop contrast in that the toddlers in the second age group used both of the acoustic parameters.

Table 5.14. The output of the mixed-effects logistic regression for the group (2;7–3;0).

Random effects:				
Groups	Name	Variance	Std.Dev.	
child_id	(Intercept)	4e-14	2e-07	
Number of obs: 138, groups: child_id, 13				
Fixed effects:				
	Estimate	Std.Error	z value	Pr (>|t|)
(Intercept)	-2.730643	1.017256	-2.684	0.00727 **
VOT	0.011614	0.005817	1.996	0.0459 *
F0	0.005164	0.002523	2.047	0.04067 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

The third group, aged 3;1 to 3;6, shows that the role of VOT tends to be weakened ($p < 0.1$), while F0 appears to be an effective predictor ($p < 0.001$), in relation to the production of /k^h/. This phonetic differentiation implies that the toddlers began to recognize the fact that the most distinctive phonetic difference between velar lenis and velar aspirated stops should be their F0 values. It can also be assumed that in their perceptual space, the phonemic categories for /k/ and for /k^h/ have been built in the dimension of F0, so the appropriate phonetic differentiation could be made. Recalling the labial stop production pattern of the same toddlers, they used F0 as a crucial parameter to determine aspirated stop /p^h/. This means that F0 development regarding the articulatory distinction between the lenis and aspirated stops has occurred before or during the period of 3;1 to 3;6.

Table 5.15. The output of the mixed-effects logistic regression for the group (3;1–3;6).

Random effects:				
Groups	Name	Variance	Std.Dev.	
child_id	(Intercept)	0	0	
Number of obs: 149, groups: child_id, 17				
Fixed effects:				
	Estimate	Std.Error	z value	Pr (>|t|)
(Intercept)	-4.524381	1.09171	-4.144	3.41e-05 ***
VOT	0.008583	0.004771	1.799	0.072 .
F0	0.011971	0.003022	3.962	7.44e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The oldest age group was expected to have clearer F0 differentiation for /k^h/ as observed in Table 5.16. F0 is significantly related to the phonation types ($p < 0.001$), while VOT does not play a crucial role anymore ($p > 0.1$). The fixed effects coefficient for *F0* increases with age, which implies that in the production of the older children, the effects of F0 would be more emphasized.

Table 5.16. The output of the mixed-effects logistic regression for the group (3;7–3;11).

Random effects:				
Groups	Name	Variance	Std.Dev.	
child_id	(Intercept)	0	0	
Number of obs: 114, groups: child_id, 14				
Fixed effects:				
	Estimate	Std.Error	z value	Pr (>|t|)
(Intercept)	-4.181148	1.315384	-3.179	0.00148 **
VOT	-0.00552	0.005652	-0.977	0.328741
F0	0.014577	0.004115	3.543	0.000396 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Figure 5.12 illustrates how the role of F0 or VOT changes depending on age group. With increased age, the relationship between F0 and the production of aspirated stops becomes stronger. Generally the role of VOT seems minimal or redundant, but it is interesting to find that the second youngest age group, aged 2;7 to 3;0, showed a strong relationship between VOT values and the phonetic distinction between aspirated and lenis stops. The children showed that both VOT and F0 phonetically characterize their production of /p^h/ and /k^h/ even though the effect of F0 seems much weaker. When F0 does not solely function enough to distinguish those contrasts, it seems that VOT differentiation preferably takes place, as shown in the production of the second age group. However, as F0 develops to the extent to operate in articulatory discrimination, it is found that the role of VOT becomes ineffective. As children grow, F0 begins to work as a crucial acoustic parameter to determine aspirated and lenis stops and the VOT differentiation diminishes, which can be shown in the production of the two oldest age groups (3;1–3;11). This indicates that 3 years of age is a critical period for the development of F0 as an acoustic cue for Korean stop contrasts.

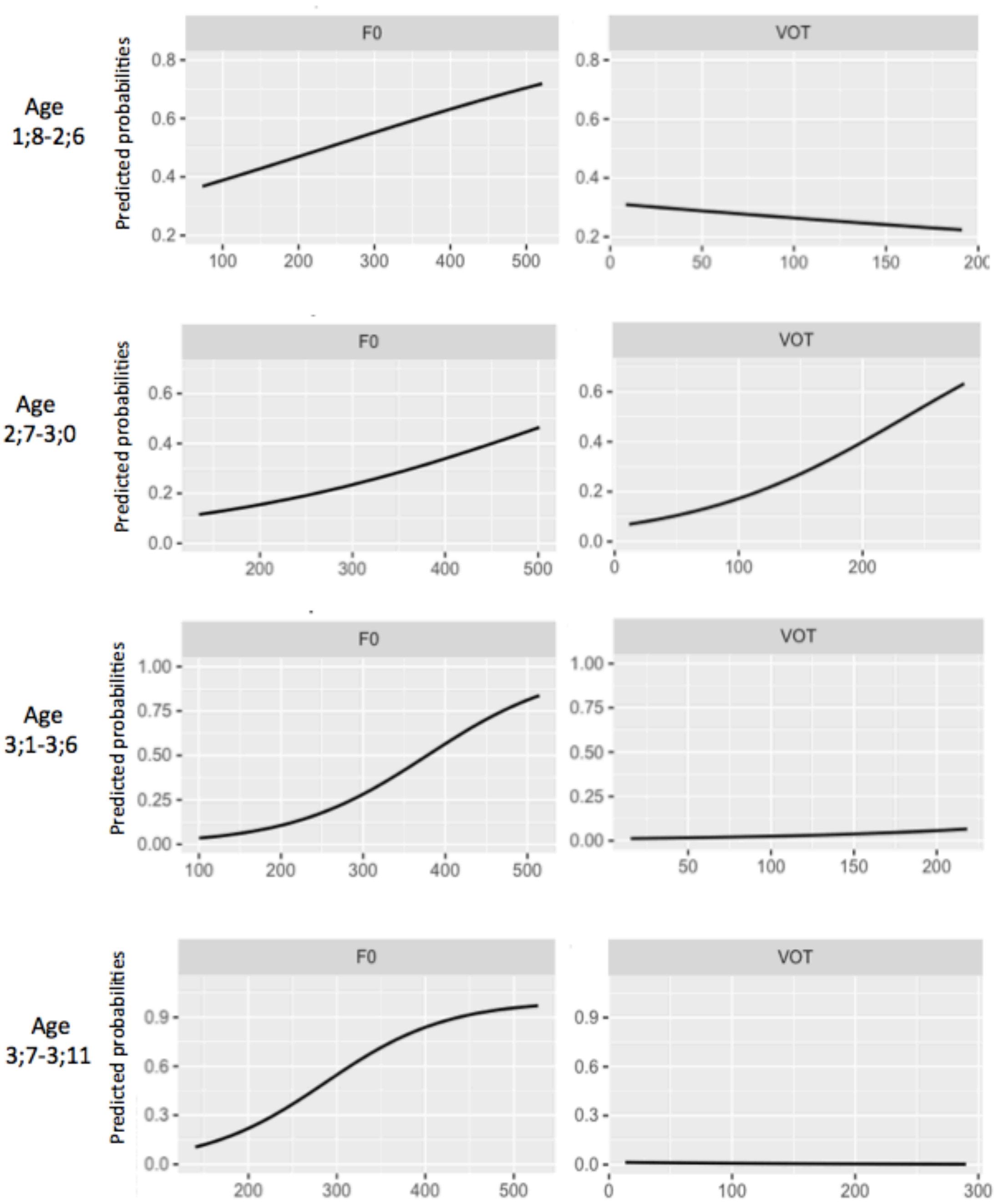

Figure 5.12. Predicted probabilities for /k^h/ with F0 and VOT difference in four child age groups. (Top row: 1;8-2;6, second row: 2;7-3;0, third row: 3;1- 3;6, bottom row: 3;7-3;11)

5.3. Comparative analysis: Production vs. Perception

In order to understand the development mechanism of native acoustic parameters, it is pertinent to investigate what interaction exists between children's productions and perceptual ability. This study attempted to uncover how child speakers are able to distinguish F0 ranges for the two phonemic categories for lenis and aspirated stops over age. The interactive development of phonemic categorization between articulatory distinction and perceptual identification was analyzed through an effective comparison among the four age groups of child speakers.

We recall that the perception experiment with the synthesized stimuli revealed that children's age is not directly related to their perceptual identification patterns, but F0 values of the stimuli significantly affect them. According to the relative F0 differences from the lenis stops, children's responses varied, as seen in Figure 5.13.

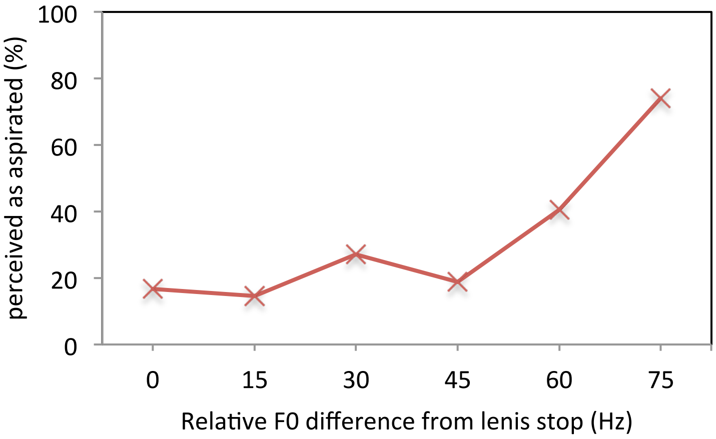

Figure 5.13. Children's responses to the relative F0 differences from lenis stops.

As the F0 difference increases compared to the standard lenis stop, the possibility of being identified as an aspirated stop would be higher. The 0 to 45 Hz difference from the lenis stop seems to have similar effects on children's perceptual identification of a stop, since most of the children judged these as sounding like lenis stops. The children's perceptual identification patterns indicate that perceptual sensitivity responds to the relatively large amount of F0 difference over 60 Hz and that this amount of F0 difference causes them to recognize it as aspirated. The stimuli with a 75 Hz difference from the lenis stop tended to be identified as aspirated, by 74% of the children.

In order to compare to the children's lenis-aspirated stop distinction in their production data, the F0 values at vowel onset after the target stop sounds were estimated, and the average F0 differences between the two stop categories were calculated as in Table 5.17. Since aspirated stops usually have higher F0 values, the F0 differences from aspirated to lenis stops were calculated.

In the case of /k^h/-/k/, the F0 differences apparently become larger with age, but in the case of /p^h/-/p/, the three younger age groups have similar F0 differences around 40 Hz. Since the four age groups showed small F0 differences in the case of /p^h/-/p/, it is consistently suggested that age difference is not always related to a significant F0 difference. It can be interpreted that even younger children who are included in the youngest age group (2;0-2;6) were able to produce a substantial F0 difference (around 40 Hz) for the phonetic distinction between /p/ and /p^h/, though a 40 Hz difference is not as much as the adult speakers have in their productions. After 3;6, the children come to have much more substantial F0 differences up to 63.5 Hz, which is almost identical with the amount of phonetic difference in adult speech.

There are some differences in F0 differentiation between the two sets; the labial contrasts have a relatively substantial difference around 40 Hz even in the youngest group, while in the case of /k/-/k^h/, the two youngest age groups show very poor F0 differentiation. However, the children over 3;0 showed quite substantial F0 differences between the velar contrasts, as much as the adult speakers.

Table 5.17. The mean F0 differences (Hz) in children's productions by age group.

Age group	/p^h/−/p/	/k^h/−/k/	Mean F0 diff.
2;0−2;6	40.0	24.5	32.2
2;7−3;0	39.0	28.2	33.6
3;1−3;6	40.9	65.2	53.0
3;7−3;11	63.5	70.9	67.2
Adults	63.6	66.6	65.1

The relative differences in F0 between aspirated and lenis stops produced by the children are illustrated with standard errors in Figure 5.14. As mentioned earlier, relative differences in F0 between two stop manners come to increase to some extent. The youngest age group shows no substantial F0 difference between their actual productions of lenis and aspirated stops. However, the oldest age group shows almost similar F0 differences to adult speakers. Another finding to note is that there is some difference in the amount of the F0 differences for velar and labial stop contrasts. The phonetic distinction between /k^h/ and /k/ was made with a 65 to 70 Hz difference in F0 by relatively older children, which is a higher F0 range compared to the distinction in the labial stop contrast /p^h/-/p/. Since adult speakers' productions show that the two different POAs share the same relative F0 differences between the different phonation types, these phonetic differences are not considered intrinsic, i.e., not physiologically caused by different POAs. In children's production, the discrepancy of F0 differences required for the distinction for the labial and velar stop contrasts indicates that identical phonation types do not share identical F0 ranges, but depending on the phoneme, children's perceptual categorization in the dimension of F0 can differ.

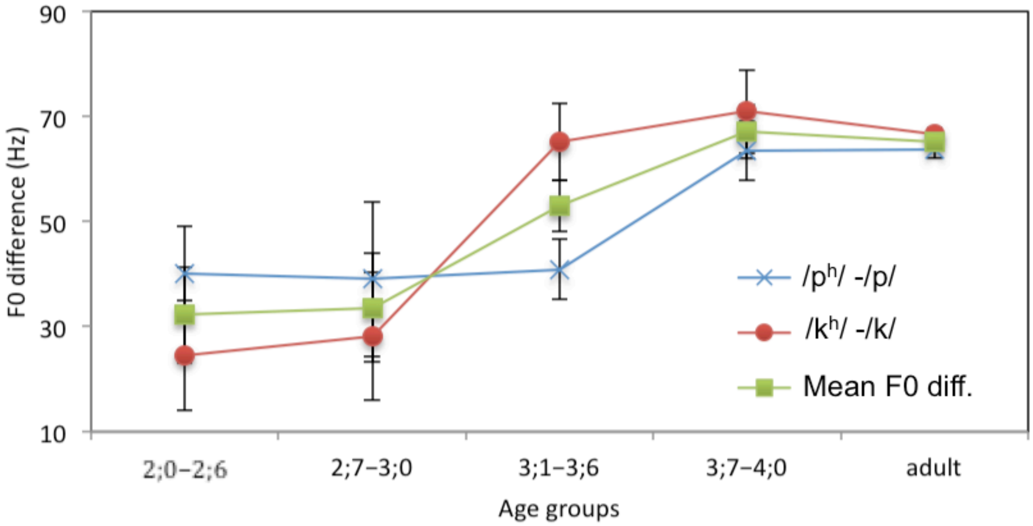

Figure 5.14. F0 difference in the productions of the four child age groups and female adult speakers. Error bars represent standard errors.

The most unbalanced F0 differentiation between the labial and velar stop contrasts is found in the third age group's production. According to Figure 5.14, before 3 years (3;0), the F0 differentiation between /k^h/ and /k/ seems minimal, but after 3 years, an exaggerated F0 differentiation is consistently observed, while F0 differences between /p/

and /p^{h}/ do not reach the extent of the difference in the velar pairs. This reversed F0 differentiation pattern around 3 years indicates that the children in the third age group (3;1-3;6) have experienced a certain linguistic development that affects their phonetic distinction patterns.[12] As statistically analyzed in Sections 5.1.2 and 5.2.2, in the production of the third age group, there was no significant VOT contribution to produce a phonetic distinction between lenis and aspirated stops, but F0 differences between the two phonation types were significantly large. Therefore, F0 differences between /p^{h}/ and /p/ in the production of the third age group did not show a substantial increase, but it is still considered a statistically significant phonetic difference for distinguishing stop contrasts while VOT differentiation did not work.

Another possibility to account for these different F0 ranges between labial and velar groups would be that the older children tend to exaggerate their pronunciations, especially when they articulate velar stop contrasts. The acquisition of velar stop contrasts is relatively delayed compared to the acquisition of labial contrasts, so the exaggerated F0 differences between /k^{h}/ and /k/ could represent immature production of the velar pair. Children’s effort to have a substantial F0 difference could be represented as this exaggerated F0 difference. It is also possible that lenis /k/ was replaced with fortis /k’/, producing the large amount of F0 difference. Recall that immature linguistic performance was found in the perception of /k/ and /k^{h}/ as well. Development of the

---

[12] Since the third age group’s productions of labial and velar stop contrasts are different, we decided to test the same aged children with synthesized /t/ stimuli. With data of the alveolar stop contrasts, it could be helpful to understand the motivation for their mysterious production patterns that are represented as an exaggerated F0 differentiation only in the velar stop contrasts. It was difficult to find monosyllabic natural minimal pairs with /t/ and /t^{h}/, so proper nouns (/taŋ/ and /t^{h}aŋ/) with unfamiliar objects were used for this experiment. To compare with the results of the experiment with /t/, see Appendix.

perceptual side is responsible for proficiency in production, so it is quite understandable that the children's articulatory distinction between /k/ and /k^h/ has not yet completed, but over time the degree of F0 difference between those two stop categories will become stabilized around 60 Hz as observed in the case of labial stops.

Recall that not every child could provide stable and consistent responses to the synthesized stimuli, but with age, the ratio of stable responses to all obtained responses increases. In order to investigate the phonemic categorizing mechanism on the perceptual side, the production and perceptual patterns of the stable responders only were analyzed.

Figure 5.16 represents the mean F0 difference that was required to identify stimuli as aspirated stops in the stable responders' perceptual identification. The obtained F0 values that are considered a perceptual threshold for aspirated stops were calculated based on their responses. The values were calculated at the midpoint between the lowest F0 value of the stimulus that was identified as aspirated and the highest F0 values of the stimulus that was identified as lenis, as shown in Figure 5.15. For example, if a child responded '/p/, /p/, /p/, /p/, /p^h/, /p^h/' to '200, 215, 230, 245, 260, 275 Hz' stimuli respectively, then the minimum F0 difference for categorizing /p^h/ is calculated as 67.5 Hz (= (the midpoint of 260 and 275, 267.5 Hz) – (lenis /p/ F0 value, 200Hz)) in this case.

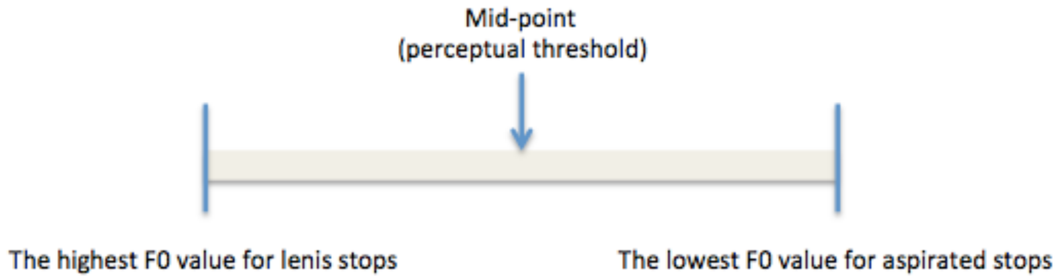

Figure 5.15. The calculation of perceptual thresholds for aspirated stops in children's perception.

In Figure 5.16, the relative F0 differences between aspirated and lenis stops vary depending on each individual and according to age group; there is no criteria consistently found to determine aspirated stops. The consistently observed phenomenon is that most of the children require a larger amount of phonetic differentiation in F0 between lenis and aspirated stops in the case of velar stop contrasts; however, the third age group's (3;1–3;6) perceptual thresholds for aspirated stops were identical for labial and velar stops, which is around a 45 Hz difference. Through this perceptual pattern, it can be said that the toddlers' perceptual mapping is still in process, since the required F0 differences for identifying /k^h/ are not stabilized even in the case of the oldest age group. In addition, the pattern that the two different sets of the lenis-aspirated contrast do not need a similar degree of relative F0 difference indicates that in children's perception, the velar aspirated stop is more characterized by a larger amount of F0 difference from its lenis counterpart compared to the labial set. It also means that perceptual F0 development depends on the phoneme rather than the phonation type itself in early stages of language development.

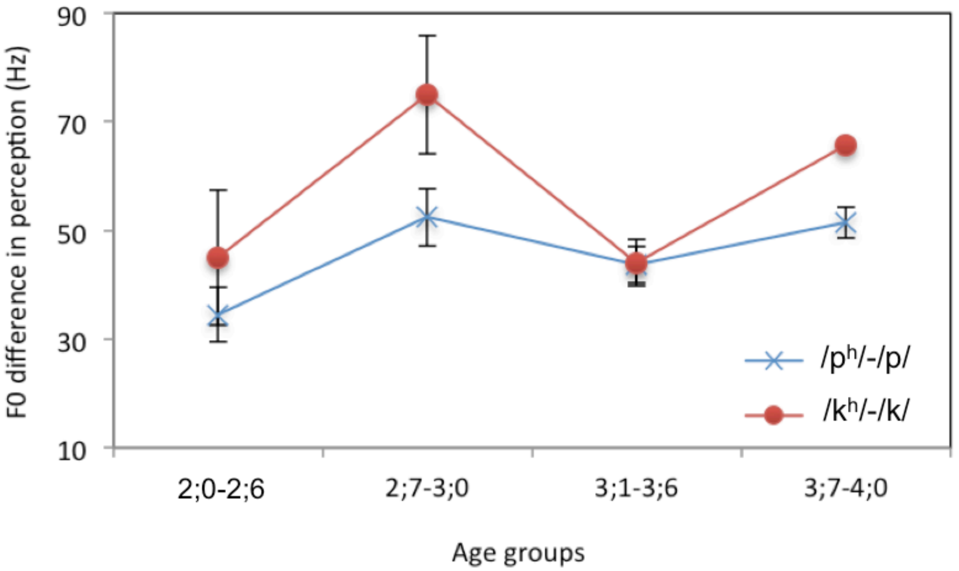

Figure 5.16. F0 differences in perception of stable responders. Error bars represent standard errors.

The production patterns of the same stable responder child participants are illustrated as in Figure 5.17. For comparison to the results of every child participant, the production patterns of the stable responders and all children are represented on the same plot. This is because the numbers of stable responders are relatively small to generalize. The calculated F0 differences between aspirated and lenis stops in the production of the stable responders are very similar across the different POAs, showing that over age, relative F0 differences keep increasing. In the production of the youngest age group, the children showed subtle F0 differences between /k^h/ and /k/, while they showed a more robust phonetic difference between /p^h/ and /p/ (even though the amount of F0 difference is only around 16 Hz). Considering that in their productions, there are large standard errors calculated, the phonetic differentiation with F0 in production seems almost immature. Until the age of 3 years, the amount of F0 difference between lenis and aspirated stops seems minimal with large standard errors, but after that age, the children's production pattern changes sharply; the phonetic differentiation for /k^{h}/ and /k/ is somewhat exaggerated such that the estimated F0 difference between them is larger than that between /p^{h}/ and /p/.

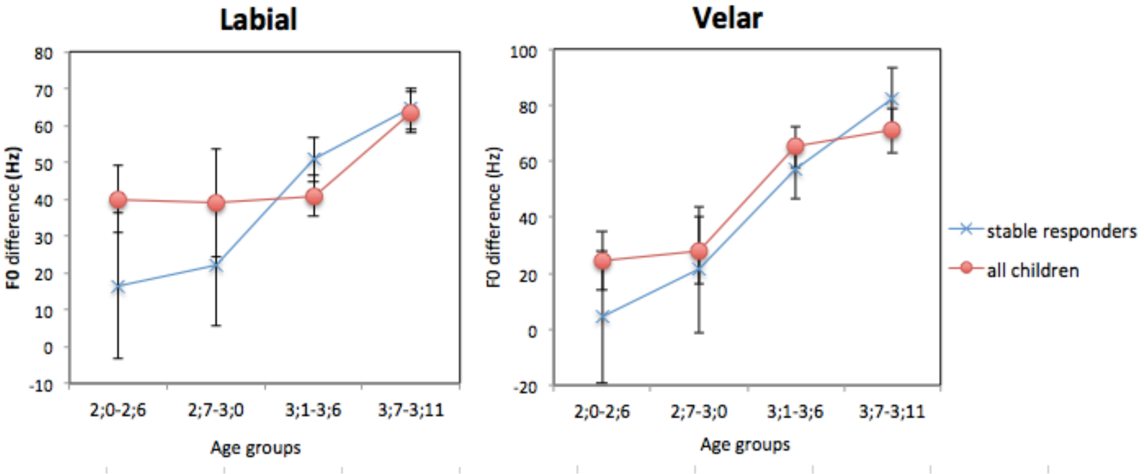

Figure 5.17. The F0 differences between productions of lenis and aspirated stops in labial (left panel) and velar (right panel) positions, as produced by stable responders and all child participants. Error bars represent standard errors.

When children's articulatory skills regarding F0 differentiation are immature, their phonetic distinctions are not great enough to be clearly represented in their actual pronunciations. It seems that they need more substantial phonetic differences even in their perception. Therefore, there is an extreme discrepancy between production and perception in the case of younger children. It is notable that perceptual distinction of phonological contrasts is acquired prior to the development of production, and the children's perceptual thresholds for aspirated stops experience phonetic adjustment as they receive linguistic input, resulting in stabilization at some phonetic criterion. This process makes even the younger children have nearly equivalent perceptual F0 thresholds to the older children.

As seen in Figure 5.16, in perception, the children in the oldest age group only needed 50 Hz and 65 Hz to categorize, /p^h/ and /k^h/ respectively, but a much larger amount of relative F0 differences is observed in their actual productions; 65 Hz for /p^h/ and 81 Hz for /k^h/. This disjoint in F0 differences between production and perception indicates that perceptually the children are more sensitive, so to some extent, they could recognize crucial acoustic characteristics to appropriately categorize perceptual input. Children's articulatory development represented as phonetic differentiation does not correspond with their perceptual development in this regard until the age of 4 years.

The different degrees of phonetic differences in production and perception have been suggested in the framework of the Theory of Adaptive Dispersion (TAD), in which phonetic parameters are build at the compromising point of the two different needs in production and perception (Liljencrantz & Lindblom, 1972; Lindblom, 1990; Lindblom & Engstrand, 1989). The F0 differences calculated from adult speakers' production and perception could be taken as evidence in support of the framework of the TAD, in that it is found that perceptually sufficient phonetic differences and articulatory economy compromise at a certain point in the F0 dimension, as shown in Figure 5.18.

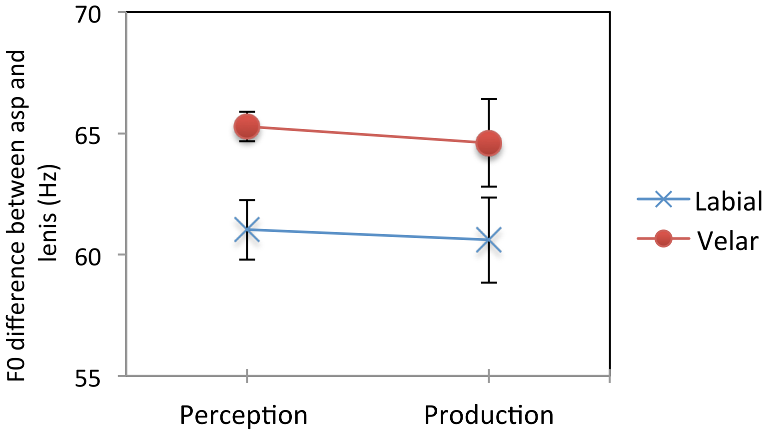

Figure 5.18. F0 differences between aspirated and lenis stops in adult speakers' production and perception. Error bars represent standard errors.

In the perception of adult speakers, an F0 difference of as much as 65 Hz was required for the distinction between lenis and aspirated stops, while a similar amount of F0 differences was estimated in their articulatory distinction. This indicates that phonemic boundaries in their acoustic space have been established to consistently work to map the phonetic input onto the appropriate phonemic categories, so consistent distinction between lenis and aspirated stops can be found in their production and perception.

In contrast, the estimated F0 differences in the child participants' production and perceptual discrimination show a different pattern in Figures 5.19 and 5.20. As previously mentioned, the immature production pattern by the relatively younger children is represented as subtle F0 change with large standard errors depending on the stop category. As seen in Figure 5.19, the two younger groups show a phonetic discrepancy between their production and perception, but after 3 years (3;0), the pattern is reversed so that the F0 difference in production is larger than that in perception. The F0 difference in perception was the highest in the second youngest group, but in the third age group, it stabilized around 50 Hz. The exaggerated high F0 values in production are, however, still observed in the two oldest age groups.

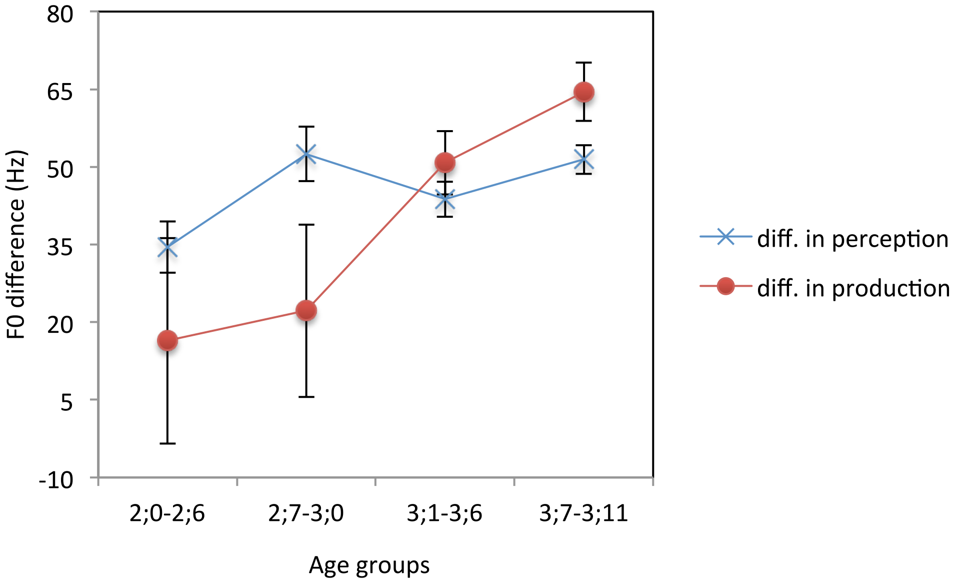

Figure 5.19. Comparison of F0 differences in production and perception of labial stop contrast /p^h/-/p/ of stable responders. Error bars represent standard errors.

In the case of velar stop contrasts in Figure 5.20, very similar patterns to the labial contrast are observed. The perceptual identification pattern of the second youngest group showed a peak in their perceptual thresholds for aspirated /k^h/. After the age of 3 years (3;0), the perceptual thresholds lowered to around 45 Hz but were soon recovered at 65 Hz at the age of 3;7-3;11. In both cases, the production of the older children shows a larger amount of F0 difference between lenis and aspirated stops while their perceptual thresholds are lowered.

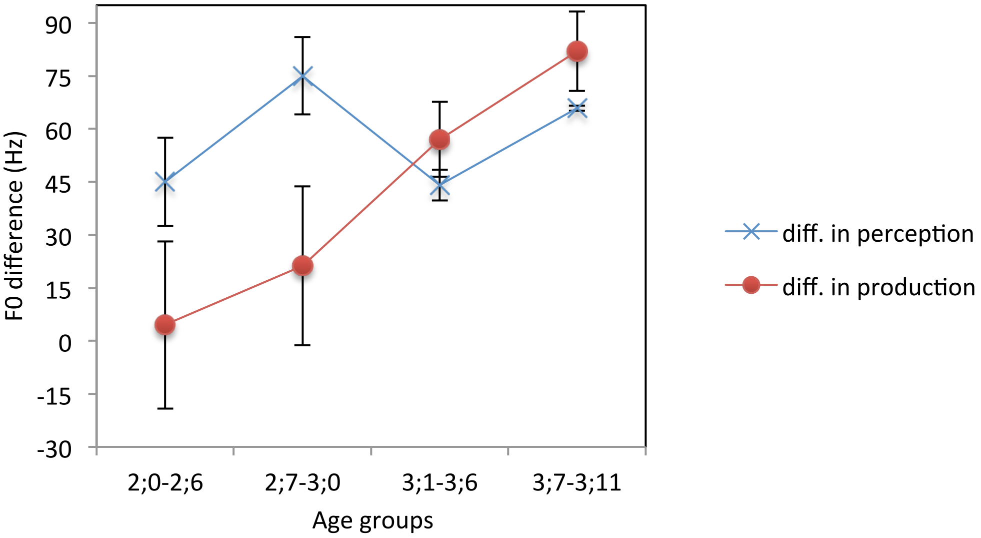

Figure 5.20. Comparison between F0 differences in production and perception of velar stop contrast /k^h/-/k/ of stable responders. Error bars represent standard errors.

Through the interaction between production and perception of stop contrasts, the commonly observed phenomenon is that there is a reversal in production and perceptual F0 differences around 3 years. The children's phonetic differentiation patterns in production and perception are reversed, since they tended to hyperarticulate stop sounds while they were perceptually more sensitive, so a smaller amount of F0 differences in their perception was found. This reversed pattern indicates that the language development in target-aged toddlers is still in progress, decreasing their perceptual sensitivity and making phonetic adjustments for their native phonological contrasts.

Thus, the age of 3 years seems to be a crucial point in critical F0 development in both the production and perception systems. Around 3 years, children's articulatory distinctions could reach the perceived acoustic differences. This process must occur through a phonemic categorization in the acoustic dimension of F0 in their perception. In this multi-parametric acoustic space, VOT and F0 develop enough to promote the appropriate phonemic categorization with those acoustic parameters, and this process is stabilized to operate properly for production of stop contrasts.

Therefore, it is remarkable that perception and production systems are interrelated and that perceptual linguistic development motivates acoustic implementation. Considering this fact, this study attempted to investigate how young children would respond to relative phonetic differences while decreasing their perceptual sensitivity with age. Through this analysis, it is possible to understand how the formation of the phonemic categories is related to phonetic distinctions in production. This interactive developmental process is illustrated in the process of F0 acquisition regarding the stop contrasts between lenis and aspirated stops.

5.4. General discussion

As VOT dominantly functions quite exclusively for distinguishing fortis stops, it is important to acquire F0 as the determinant cue to phonetically differentiate lenis stops from aspirated stops (Kang, 2014; Kim, 2008; Kong *et al.*, 2011). This chapter is devoted to investigating how the development of F0 affects phonemic categorization for the lenis-aspirated stop contrast in Korean, and how this developmental process impacts the perceptual and articulatory distinction of the stop contrast.

The comparison between adult speech and perception supports the main factors in the framework of the Theory of Adaptive Dispersion in that their perceptual mapping and production of stops were consistently identical. The adult speakers showed categorical perception in the perception test with an F0 continuum, since they consistently perceived aspirated stops only at a certain range of F0. Relative F0 differences from the lenis counterparts were consistently estimated in their production and perception.

On the other hand, the children's stop perception and production showed that phonemic category formation is in progress in the acoustic dimension of F0, which directly affects the discriminatory ability for lenis-aspirated contrast. The two different sets of lenis and aspirated stops have somewhat different phonetic details but also showed a similar development trajectory in acquisition of F0 with respect to distinction between them. For example, it seems that in children's speech processing development, the critical acoustic development occurs around 3 years of age. In the /p/-/p^h/ set, the significant F0 differentiation emerges in stop discrimination while VOT differentiation is lost after 3 years of age. In the /k/-/k^h/ set, the role of F0 becomes determinant but VOT differentiation decreases after 3 years of age.

It is remarkable that these articulatory distinction skills would be perceptually driven since in perception, around 3 years of age, the children's perceptual F0 threshold for aspirated stops was becoming similar to that of adult speakers. Before 3 years, the children were more perceptually sensitive, so their perceptual F0 thresholds for aspirated stops were lower compared to the adult speakers' results. This indicates that their phonemic reorganization in the acoustic dimension of F0 is still in progress and needs further phonetic adjustment to stabilize. In this process, young children need to stabilize the sufficient contrastive distinctiveness for lenis and aspirated stops. Establishing phonemic boundaries for lenis and aspirated stops, they were able to respond consistently to the F0 continuum after the age of 3;0. After 3 years of age, children depend solely on F0 differences for phonemic categorization of lenis and aspirated stops. This perceptually driven development process leads to more mature acoustic implementation, so their articulatory distinctions showed quite consistent phonetic realization with F0.

The one thing to note is that children's articulatory distinctions were exaggerated so that relative F0 differences between the production of aspirated and lenis stops were larger than the F0 differences estimated in their perception. This hyperarticulation with respect to F0 could be an indication of the children's immature articulatory skills. Since phonemic categorization in the acoustic dimension of F0 has not been completed in their perceptual space, their production could be affected. The exaggerated production pattern implies that their F0 differentiation has not stabilized but is still developing.

As a result, the process of acquiring F0 provides of experimental evidence for theories about the phonetic function on the phonology such as the Perceptual Assimilation Model (Best, 1993; 1995; Best, McRoberts, & Goodell, 2001; Best, McRoberts, & Sithole, 1988) and the Theory of Adaptive Dispersion (Liljencrantz & Lindblom, 1972; Lindblom, 1990; Lindblom & Engstrand, 1989). The main factor of these theoretical models suggests that phonemic categorization is motivated by perception of phonological features in the phonetic input. With the concept of this active perceptual pressure in phonemic processing, the cleft found between the children's production and perception of stop contrasts can be explained at least in part. Toddlers' linguistic development proceeds with phonetic input to form phonemic categories in the

F0 dimension. This study attempted to capture the developmental trajectory in the dimension of F0 during the period of 2 to 4 years of age, and revealed that perceptually driven acoustic development is directly linked to articulatory skills. Therefore, the observed articulatory-perceptual discrepancy in the stop contrast itself can be a piece of empirical evidence for the acquisition of phonological contrasts on theoretical grounds.

Chapter 6
Summary and concluding remarks

6.1. Main findings

This study aims to investigate how Korean young children acquire and use F0 in their perception and production to make a phonemic distinction between lenis and aspirated stops through experiments. As tonal contrasts have developed in standard Korean, the role of F0 becomes determinant for distinguishing between stop contrasts, and it is worth investigating how this tonogenetic process can be related to young children's acquisition of F0.

The present study conducted the two different experiments with 20- to 47-month-old toddlers. The production experiment revealed that VOT is acquired prior to F0 at around 2 years of age or earlier, resulting in the earlier acquisition of fortis stops than the acquisition of lenis or aspirated stops. The phonetic definition of fortis stops is significantly short VOT values in the toddlers' and adults' productions. On the other hand, the phonetic distinction between lenis and aspirated stops tends to be delayed mainly because the development of F0 occurs later than VOT. Therefore, the results consistently indicate that significant distinction between lenis and aspirated stops requires a phonetic difference in the dimension of F0, which seems to develop between 2 to 4 years of age. In addition, the production of lenis stops by the children was the most immature, so the statistical analysis shows that the acquisition of lenis stops come latest and is directly related to more specific and accurate F0 differentiation from aspirated stops.

The perception experiments indicate that the perceptual identification of natural stimuli of fortis stops depends on VOT differences, while F0 differences are dominant in affecting the perception of aspirated stops. This perceptual pattern is consistent with the production pattern in that VOT is the most relevant phonetic parameter for the identification of fortis stops. The perception of lenis and aspirated stops is directly involved in appropriate F0 differentiation, and lenis stops had the lowest accuracy level in perceptual identification. Stop category was relatively accurately identified when VOT differences were emphasized, while it was less accurately perceived when F0 had to be the main cue to distinguish for it. Generally, young children aged 24 months to 47 months successfully perceived natural sounds of phonological contrasts.

For investigating the development of F0 in perceptual acoustic space, sound stimuli synthesized to have various F0 onsets were provided to the same children. This made it possible to determine how their perceptual sensitivity to F0 changes with age. The results indicate that perceptual sensitivity to F0 decreases with age and that this perceptual attunement to F0 significantly affects children’s stop identification patterns. This means that at some phonetic standard, they began to distinguish a lenis-aspirated contrast and their phonetic boundaries regarding F0 are adequately settled. Through this reorganization of phonemic stop categories, children were able to correctly and accurately perceive two different phonation types.

Through a comparison of the output of the two experiments, it was possible to uncover the interactive F0 development between children’s production and perception systems in the process of acquiring general language competence and to assess how F0 comes to play a determining role while the role of VOT declines in distinguishing lenis and aspirated stops. Before the age of 3 years, VOT significantly functions for distinction between lenis and aspirated stops. However, after the age of 3 years when F0 has developed enough to function for phonetic distinction, the phonetic trade-off between VOT and F0 has occurred. This acquisition process allowed us to suggest that VOT and F0 complementarily function and that young children did not simultaneously use both of the acoustic parameters for discrimination between stop contrasts. It is interesting that young children’s F0 acquisition pattern recapitulates the tonogenetic sound change found in young adult speech in recent standard Korean of a diminished VOT and an amplified F0 for distinguishing lenis from aspirated stops (e.g., Kang, 2014). This acquisition pattern would be considered language-specific because it has been known that in acquisition of tone languages such as in Mandarin Chinese, lexical tonal contrasts are acquired and mastered earlier than segmental features (Li & Thompson, 1977).

Since children’s perceptual identification is more successful in earlier age compared to their production of stops, it can be suggested not only that perceptual development occurs earlier than production skills, but that it is also responsible for appropriate acoustic implementation. Interestingly, hyperarticulation of the lenis-aspirated contrast was found in some groups of older children (who were 42- to 47-months old); F0 differentiation in production was more exaggerated, so the relative F0 differences between lenis and aspirated stops, which were estimated through the perception test, tend to be smaller in perception.

The findings from the present study provide experimental evidence for major theoretical models on young children's perceptual learning such as in the Native Language Magnet Effect (NLM) and the Theory of Adaptive Dispersion (TAD). This comparative analysis between the production and perception of the Korean stop series attempted to capture the process of the establishment of appropriate phonemic categories using F0 in perceptual space. Over age, the target children showed that relative F0 differences between aspirated and lenis stops increased to a similar extent as articulatory F0 differences. Once their phonetic boundaries in F0 are established, children's language competence allows more accurate phonemic perception and production. Therefore, the process of F0 development between the ages of 2 to 4 years supports the essence of NLM in that the role of phonemic categories are crucial in distinguishing phonological contrasts. In addition, the changing roles of VOT and F0 in distinguishing between lenis and aspirated stops also can be seen as evidence for TAD, since children try to maintain valid perceptual contrast for aspirated and lenis stops using F0 instead of VOT over age.

6.2. Limitations and future direction

As the main findings help to capture a trajectory of F0 development, this study provides empirical evidence for the acquisition of native phonological contrasts by toddlers. This work makes a distinguishable contribution to the literature in that this study attempts to draw a clear correlation line between toddlers' age and linguistic competence that is represented as significant phonetic distinction among stop contrasts. However, it is also true that it was hard to unify developmental patterns across the target ages from 20 months to 47 months, and mysterious perceptual patterns were also found. Recall that the third age group (3;1 – 3;6) showed inverted developmental patterns in their production and perception of stop contrasts. In order to address this issue, future works should focus on the second and the third age groups, 2;7 to 3;6, since it seems that an important stage of F0 development occurs during this period.

In spite of the fact that this study attempted to show a comparative analysis between the same toddlers' production and perception, it was difficult to obtain a large amount of stable responses from the youngest age group, 2;0 to 2;6, so the analysis of this group was inevitably generalized to some extent. Increasing the total number of child participants would increase the probability of getting valid responses, which would produce more reliable results.

Another challenge in the analysis was the limited F0 values at vowel onset in the second perception experiment. The maximum relative difference in F0 was 75 Hz in the actual experiment, but it was discovered that greater F0 differences between lenis and aspirated stops should have been tested in order to clearly observe categorical perception.

Of course, the number of stimuli should be carefully restricted in order to use for the experiment with young aged toddlers since their concentration span is usually very short.

The most important thing to investigate further in relation to this research would be about the effect of phonetic properties of children's linguistic input on the phonetic trade-off between VOT and F0, which was found around the age of 3 years. As previously pointed out, young children's F0 developmental pattern quite resembles the tonogenetic sound change in progress in young adult speech in standard Korean: VOT differentiation declines while F0 differentiation is enhanced. During this process of tonogenesis, it is expected that inconsistent adult speech in terms of VOT and F0 differentiation might affect young children's language development. Therefore, what acoustic characteristics in children's perceptual input cause such F0 enhancement in distinction between lenis and aspirated stops should be studied in relation to a language-specific F0 acquisition pattern.

APPENDIX

Perception Experiment with synthesized /t/ stimuli

Table A.1. Child participant information.

Age (year;month)	Male	Female	Total
3;0	1	0	1
3;1	1	0	1
3;2	0	1	1
3;3	3	4	7
3;4	1	1	2
Total	6	6	12

Table A.2. Proper nouns (nonce words) used for the perception experiment.

Target phoneme	Word
t	/taŋ/
t^h	/t^haŋ/

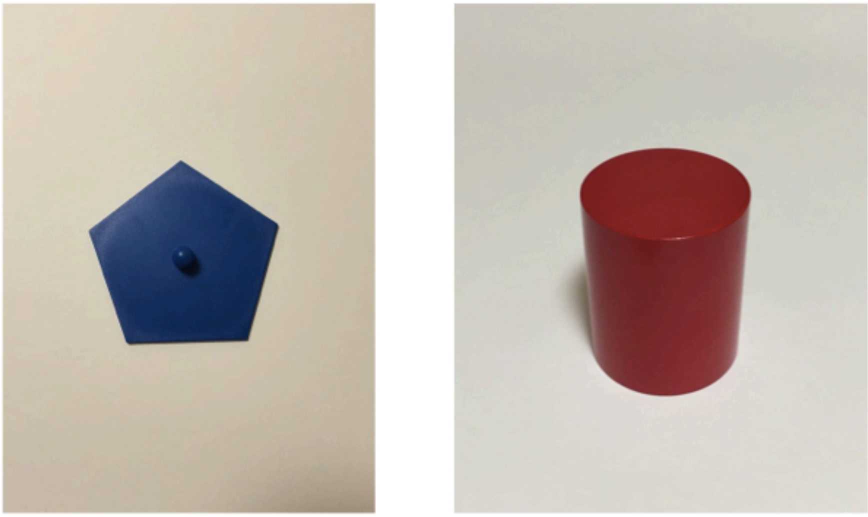

Figure A.1. Objects used for '/taŋ/' (left) and '/t^haŋ/' (right).

Table A.3. Word list for production test.

Target phoneme	Word	English gloss
t	/tak/	hen
t^h	/t^hajo/	Tayo (famous character)

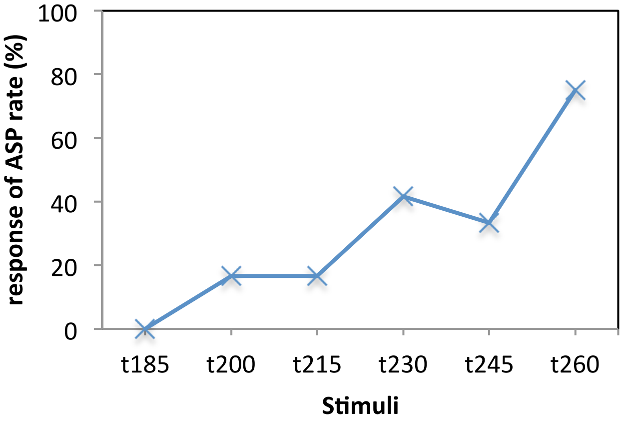

Figure A.2. The change of response rate when perceived as aspirated depending on F0 of alveolar stimuli. The numbers following ‘t’ denote the manipulated F0 values at vowel onset.

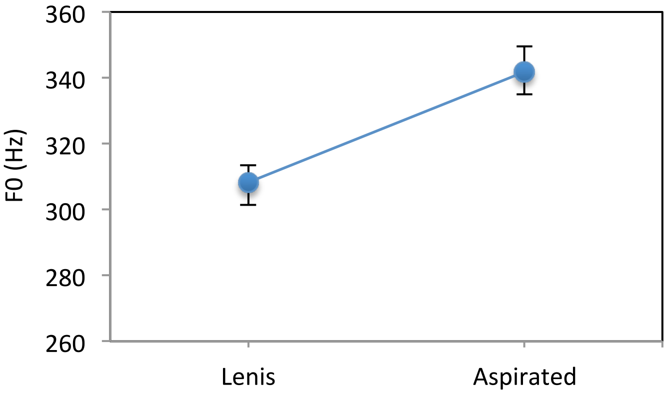

Figure A.3. In production of real words with an initial /t/ and /t^h/, F0 difference between lenis and aspirated. Error bars represent standard errors.

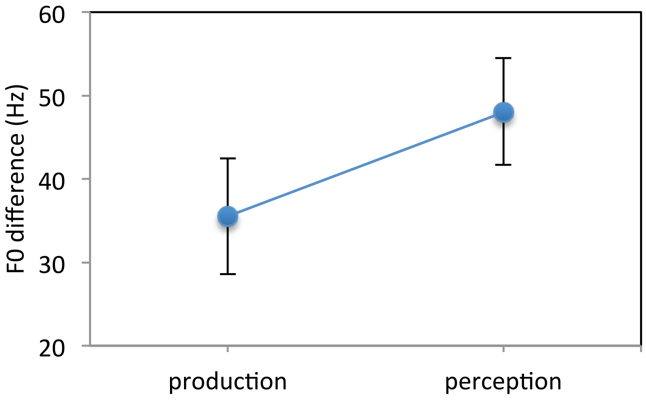

Figure A.4. F0 difference between lenis and aspirated in toddlers' production and perception. Error bars represent standard errors.

BIBLIOGRAPHY

Ahn, H. (1999). Post-release phonatory processes in English and Korean: Acoustic correlates and implications for Korean phonology. Ph.D. dissertation, University of Texas at Austin, Austin, TX.

Ahn, S. -C., & Iverson, G. K. (2004). Dimensions in Korean laryngeal phonology. *Journal of East Asian Linguistics,* 13, 345-379.

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, M. M., & Dyson, A.T. (1998). The Acquisition of Arabic Consonants. *Language, and Hearing Research*, 41, 642-653

Avery, P., & Idsardi, W. J. (2001). Laryngeal dimensions, completion and enhancement. In A. Hall, & Kleinhanz (Eds.) *Studies in Distinctive Feature Theory* (pp. 41-70). Berlin: Mouton de Gruyter.

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

Bates, D. M., Maechler, M., & Bolker, B. (2011). *Lme4 package.*

Bates, D., Maechler, M., Bolker,B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Best, C. (1993). Emergence of language-specific constraints in perception of non-native speech: a window on early phonological development. In de Boysson-Bardies, de Schonene, Jusczyk, MacNeilage, & Morton (Eds.) *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 289-304). Dordrecht: Kluwer.

Best, C. (1995). A direct realist view of cross-language speech perception. In Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp.171-204). Timonium: York.

Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners′ native phonological system. *Journal of the Acoustical Society of America*, 109, 775-794.

Best, C., McRoberts, G., & Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance,* 14, 345-360.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot international* 5(9/10), 341-345.

Boersma, P., & Weenink, D. (2015). *Praat: Doing phonetics by computer* [computer program]. Version 5.4.18. (http://www.praat.org/).

Bortolini, U., Zrnarich, C., Fior, R., & Bonifacio, S. (1995). Word-initial voicing in the productions of stops in normal and preterm Italian infants. *International Journal of Pediatric Otorhinolaryngology,* 31, 191-206.

Cho, T. (1996). *Vowel correlates to consonant phonation: an acoustic-perceptual study of Korean obstruents*. M.A. thesis, University of Texas at Arlington. Arlington. TX.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of phonetics*, 30(2), 193-228.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155-190.

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

Choi, H. (2002). Acoustic cues for the korean stop contrast: Dialectal variation. *ZAS Papers in Linguistics*, 28, 1-12.

Clumek, H., Barton, D., Macken, M. A., & Huntington, D. (1981). The aspiration contrast in Cantonese word-initial stops: data from children and adults. *Journal of Chinese Linguistics*, 9. 210-224.

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.

Eimas, P. D., Siqueland, E.R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171 (3968), 303–306.

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

Gandour, J., Petty, S. H., Dardarananda, R., Dechongkit, S., & Mukongoen, S. (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.

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Han, J-I. (1992). On the Korean tensed consonants and tensification. *CLS*, 28, 206-223. Chicago, IL: Chicago Linguistics Society.

Han, J-I. (1996). *The phonetics and phonology of "tense" and "plain" consonants in Korean.* Ph.D. dissertation, Cornell University, Ithaca, NY.

Han, N. (1998). *A comparative acoustic study of Korean by native Korean children and Korean-American children*. M.A. thesis, UCLA, Los Angeles, CA.

Han, M. S., & Weitzman R. S. (1970). "Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/," *Phonetica,* 22, 112–128.

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

Haudricourt, A. G. (1954). De l'origine des tons du Vietnamien. *Journal Asiatique*, 242, 69-82.

Hirose, H., Lee, C. -Y., & Ushijima, T. (1974). Laryngeal control in Korean stop production. *Journal of Phonetics*, 2, 145-152.

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

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment on the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society*

America, 25, 105-135.

Hua, Z., & Dodd, B. (2000). Phonological Acquisition of Putonghua. *Journal of Child Language*, 27, 3- 42.

Jacobson, R. (1968). *Child language, aphasia and phonological universals*. The Hague: Mouton.

Jessen M., & Roux J. C. (2002). Voice quality differences associated with stops and clicks in Xhosa, *Journal of Phonetics*, 30, 1-52.

Johnson, K. (2000). Adaptive dispersion in vowel perception. *Phonetica*, *57*(2-4), 181-188.

Johnson, K., Flemming, E. & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language,* 69, 505-528.

Johnson, W., & Reimers, P. (2010). *Patterns in child phonology*. Edinburgh: Edinburgh University Press.

Jun, S. -A. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. dissertation, Ohio State University, New York: Garland.

Jun, S. -A. (1996). Influence of microprosody on macroprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics*, 97-116.

Kagaya, R. (1974). A fiberscopic and acoustic study of Korean stops, affricates, and fricatives. *Journal of Phonetics*, 2, 161-180.

Kang, K. -S. (1998). On phonetic parameters in the acquisition of Korean obstruents: A case study. In C. Gruber, D. Higgins, K. S. Olson, & T. Wysocki (Eds.), *The Panels* (pp. 311-326). Chicago Linguistic Society, Chicago, IL.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: a corpus study. *Journal of Phonetics,* 45, 76-90.

Kang, K. H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America*, 124, 3909–3917.

Keating, P. A. (1985). Universal phonetics and the organization of grammars, In V. Fromkin (Ed.), *Phonetic Linguistics* (pp. 115-132). Academic Press.

Keating, P. A., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11, 277-290.

Kenstowicz, M., & Park, J. (2006). Laryngeal features and tone in Kyungsang Korean: a phonetic study. *Studies in Phonetics, Phonology, and Morphology*, 12(2), 247-264.

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

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain, *Language*, 82, 33-63.

Kim, C. -W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops.) *Word*, 21, 195-210.

Kim, C. -W. (1970). A theory of aspiration. *Phonetica*, 21, 107-116.

Kim, Y. -T. (1992). Study on developmental phonological processes of 2-6 year-old children in Seoul and Kyunggi regions. *Malsori*, 21, 3-24.

Kim, M. -R. C. (1994). *Acoustic characteristics of Korean stops and perception of English stop consonants.* Ph.D. dissertation, University of Wisconsin-Madison, Madison, WI.

Kim, Y. (1995). *Acoustic characteristics of Korean coronal*. M.A. thesis. University of Texas at Arlington, Arlington, TX.

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.

Kim, S. -H. (1999). Multiple acoustic cues of three-way phonemic contrast in stop consonants. *Studies in Phonetics, Phonology and Morphology*, 5, 79-103.

Kim, M. -R. (2000). *Segmental and tonal interactions in English and Korean: A phonetic and phonological study* [Ph.D. thesis]. Ann Arbor, Michigan: University of Michigan, Ann Arbor.

Kim, M. (2004). Correlation between VOT and F0 in the perception of Korean stops and affricates. In Proceedings of *the 8th international conference on spoken language* *processing* (INTERSPEECH-2004) (pp. 49–52). Jeju Island, Korea: International Speech Communication Association.

Kim, M. (2008). *Phonetic and phonological development of word-initial Korean obstruents in young Korean children*. Ph.D. dissertation, University of Washington, Seattle, WA.

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

Kim, M. -R., & Duanmu, S. (2004). Tense and lax stops in Korean. *Journal of East Asian Linguistics*, 13, 59-104.

Kim, H., Honda, K., & Maeda, S. (2005). Stroboscopic cine-MRI study on the phasing between the tongue and the larynx in Korean three-way phonation contrast. *Journal of Phonetics,* 33, 1-26.

Kim, H., Maeda, H., & Honda, K. (2010). Invariant articulatory bases of the features [tense] and [spread glottis] in Korean plosives: New stroboscopic cine-MRI data. *Journal of Phonetics*, 38, 90-108.

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

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge, *Language*, 70, 419–454.

Klatt, D. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18, 686-706.

Kochetov, A., & Kang, Y. (2016). Supralaryngeal implementation of length and laryngeal contrasts in Japanese and Korean. *Canadian Journal of Linguistics*.

Koenig, L. L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research,* 43, 1211-1228.

Kong, E. J. (2009). *The development of phonation-type contrasts in plosives: cross-linguistic perspectives.* Ph.D. dissertation, Ohio State University. Columbus, OH.

Kong, E. J., Beckman, M. E., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics*, 39, 196–211.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4, 812-22.

Kuhl, P. K. (1998). The development of speech and language. In T. J. Carew, R. Menzel & C. J. Shatz (Eds.), *Mechanistic relationships between development and learning* (pp. 53-73). New York: Wiley.

Kuhl, P. K. (2000). A new view of language acquisition. Proceedings of *the National Academy of Science*, 97, 11850-7.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831-843.

Kuhl, P. K., & Iverson, P. (1995). Chapter 4: Linguistic experience and the "Perceptual Magnet Effect". *Speech perception and linguistic experience: Issues in cross-language research*, 121-154.

Kuhl, P. K., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992), Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-8.

Kuznetsova, A., Brockhoff, P. B., & Christensen, H. B. (2013). *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models* (lmer Objects of lme4 pacage), R Package Version. Available online at: http://cran.r-project.org/web/packages/lmerTest/index/html

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

Lee, S., & Iverson, G. K. (2008). Development of stop consonants in Korean. *Korean Linguistics*, 14, 21-39.

Lee, S. A., & Iverson, G. K. (2011). Stop consonant productions of Korean English bilingual children. *Bilingualism: Language and Cognition*, 12, 275-287.

Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South

Kyungsang dialects. *Journal of the International Phonetic Association*, 42, 145–169.

Lehiste I., & Peterson, G. E. (1961). “Some basic considerations in the analysis of intonation,” *Journal of Acoustical Society of America*, 33, 419–425.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children, *Journal of Child Language*, 4, 185-199.

Liljencrantz, J., & Lindblom, B. (1972). Numerical simulations of vowel quality systems: the role of perceptual contrast. *Language*, 48, 839-862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In Ohala & Jaeger (Eds.), *Experimental Phonology* (pp.13-44). New York: Academic Press.

Lindblom B (1990): Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp.403-439). Dordrecht:Kluwer.

Lindblom, B. (1998). Systemic constraints and adaptive change in the formation of sound structures. In Hurford, Studdert-Kennedy, & Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases* (pp. 242-264). Cambridge: Cambridge University Press.

Lindblom, B. & Engstrand, O. (1989). In what sense is speech quantal? *Journal of Phonetics*, 17, 107-121.

Lisker, L., & Abramson, A. S. (1964). Cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, 10, 1-28.

Logan, J. S. (1992). *A computational analysis of young children’s lexicons*. Ph.D. dissertation, Indiana University, Bloomington, IN.

Lombardi, L. (1991). *Laryngeal Features and Laryngeal Neutralization*. Ph.D. dissertation, University of Massachusetts, Amherst, MA. [Published by Garland, New York, 1994.]

Macken, M. A., & Barton, D. (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., & Barton, D. (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, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, I. (1997). Phonetic universals. In W. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences*, (pp. 619-639). Oxford: Blackwell Publishers.

Metsala, J. L. (1997). Spoken word recognition in reading disabled children. *Journal of Educational Psychology*, 89, 159-169.

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

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39, 59–67.

Padgett, J. (2001). Contrast Dispersion and Russian Palatalization. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp.187-218). Academic Press.

Pae, S. (1994). Development of normal speech sounds (I): children aged 1;4-3;11. In K. S. P. Association (Ed.), *Understanding and treatment of speech and language disorders, 1. Treatment of children's articulation disorder*. Seoul:Kunja.

Pae, S., Chang, Y., Kwak, K., Sung, H., & Sim, H. (2004). MCDI-K referenced expressive word development of Korean children and gender differences. *Korean Journal of Communication Disorder*s, 9, 45-56.

Pan, H. S. (1994). *The voicing contrasts of Taiwanese (Amoy) initial stops: data from adults and children*. Ph.D. dissertation, Ohio State University, Columbus, OH.

Pegg, J. E., & Werker, J. F. (1994). Infant perception of an English allophone. Poster presented at the International Conference of Infant Studies, Paris, France.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation. MIT, Cambridge, MA.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* New York: Springer.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*. 20, 421-35.

R Development Core Team (2011). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. (http://www.r-project.org/)

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models (2nd edition)*. Thousand Oaks, California: Sage.

Rice, K. (2005). Liquid relationships. *Toronto Working Papers in Linguistics*, 24, 31-44.

Shin, J. (1997). *Consonantal production and coarticulation in Korean*. Ph.D. dissertation, University of London.

Shin, J. (2005). Phoneme frequency of 3 to 8 year-old Korean children. *Korean Linguistics,* 27, 163-200.

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

Smit, A. B., Hand, L., Freilinger, J., Renthal, J., & Bird, A. (1990). The Iowa articulation norms project and its Nebrask replication. *Journal or Speech and Hearing Disorders*, 55, 779-798.

Smith, N. V. (1973). *The acquisition of phonology: a case study*. Cambridge: CUP.

Smith, B. L. (1978a). Effects of place of articulation and vowel environment on voiced stop consonant production. *Glossa,* 12, 163-175

Smith, B. L. (1978b). Temporal aspects of English speech production: a developmental perspective. *Journal of Phonetics,* 6, 37-67.

Snedecor, G. W., & Cochran, W. G., (1989). *Statistical Methods*, 8th Ed. Ames: Iowa State University Press.

Snijders, T., & Bosker, R. (1999). *Multilevel Analysis*, London: Sage.

So, L. K. H., & Dodd, B. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, 22, 473-495.

Stoel-Gammon, C. (1985). Phonetic inventories 15-24 months - a longitudinal study. *Journal of Speech and Hearing Research*, 28, 505-512.

Stoel-Gammon, C. (1998). The role of babbling and phonology in early linguistic development. In A. M. Wetherby, S. F. Warren, J. Reichle (Eds.). *Transitions to prelinguistic communication: Preintentional to intentional and presymbolic to symbolic* (pp. 87–110). Baltimore, Md: Paul H. Brookes.

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, 1026-1037.

Sweeting, P. A., & Baken, R. J. (1982). Voice Onset Time in a normal-aged population. *Journal of Speech and Hearing Research,* 25, 129-134.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B*, 364, 3617-3632.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147-166.

Swingley, D., & Aslin, R. N. (2002). Lexical neghborhoods and the word-form represenations of 14-month-olds. *Psychological Science*, 13, 480-484.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71, 73-108.

Templin, M. C. (1957). Certain language skills in children: their development and interrelationships. *Institute of Child Welfare Monographs* (Vol. 26). Minneapolis: University of Minnesota Press.

Um, J. -H. (1994). Development of normal speech sounds (II): children aged 3,4, and 5 years. In K. S. P. Association (Ed.), *Understanding and treatment of speech and language disorders: 1. Treatment of children's articulation disorder*. Seoul: Kunja.

Van, R. C., & Erickson, R. L. (1996). *Speech correction: An introduction to speech pathology and audiology*. Boston: Allyn and Bacon.

Vihman, M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21, 61-82.

Vihman, M. (1996). *Phonological development: the origins of language in the child*. Cambridge, MA: Blackwell Publishing.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, 13, 286-350.

Werker, J. F. (1995). Chapter 5: Age-related changes in cross-language speech perception standing at the crossroads. *Speech perception and linguistic experience: Issues in cross-language research*, (pp. 155-169).

Werker, J. F., & Baldwin, D. A. (1992). Speech perception and lexical comprehension. Poster presented at Society for Research in Child Development. Seattle, WA. April, 1992.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological Development: Models, Research, and Implications,* (pp. 285-311). Timonium, MD: York Press.

Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.

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

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