Article Cover Page

Title: Why are Korean tense stops acquired so early: The role of acoustic properties

Eun Jong Kong[a]*, Mary E. Beckman[b], Edwards Jan[a],

* Corresponding author: ekong@wisc.edu, Tel: +1 608 262 6768, Fax: +1 608 263 0529
[a] Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, WI, 53705-2280 Wisconsin, USA
[b] Linguistics, The Ohio State University, Oxley Hall, 1712 Neil Avenue, Columbus, OH, 43210-1298, Ohio, USA

Abstract
Transcription-based studies have shown that the tense stop appears first in most children's speech among the three different types of homorganic stops (tense vs. lax vs. aspirated) despite its phonologically marked status. One might argue that this order of mastery is predicted by the short-lag Voice Onset Time (VOT) values of Korean tense stops, as in the cross-linguistically most common voiceless unaspirated stop type. However, the tense stop is also characterized by its creaky voice quality (negative H1-H2) and by a relatively high f0 value at the vowel onset, word-initially. In order to explain the observed order of mastery based on VOT, we need a more sensitive quantitative model of the role of multiple acoustic parameters. This study explores the relationship between native speakers' transcriptions/categorizations of children's stop productions and the three acoustic characteristics (VOT, H1-H2 and f0). The results showed that the VOT dominated in identifying the tense stop, suggesting that having a short lag VOT suffices for a transcriber or naïve adult listeners to identify a stop as a correct tense stop. Given the primary role of VOT, the early mastery of the Korean tense stop is not a cross-linguistically exceptional pattern after all.

Keywords
tense stop, Voice Onset Time, fundamental frequency (f0), transcription accuracy, Korean stop laryngeal contrast, phonological acquisition

Why are Korean tense stops acquired so early: The role of acoustic properties

1.0 Introduction

Since the seminal study by Lisker & Abramson (1964), VOT has been used as a successful acoustic measure in categorizing the stop categories across many languages. VOT is a continuous measure of the temporal relationship between two acoustic events that signal the onset of vocal fold vibration and the oral constriction release, but the three qualitatively different VOT relationships of "before" versus "simultaneous with" versus "well after" does capture the three most commonly attested phonation categories across languages. For example, in Lisker & Abramson's original cross-linguistic investigation of stop VOT distributions across eleven languages, the two stop phonation types in Dutch, Hungarian, Spanish, and Tamil could be classified as "true" voicing contrasts that differentiated between lead and short lag values, whereas the two stop phonation types of Cantonese and English could be classified as aspiration contrasts between the long lag VOT range for aspirated stops and the short lag VOT range for unaspirated stops. The three stop phonation types of Eastern Armenian and Thai were a simple union of these two two-way contrasts.

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

In this paper, we examine how the use of VOT and other acoustic properties helps to predict the mastery pattern for stop phonation types in Korean. Korean is anomalous in having a three-way contrast that does not use the "default" voiceless category, but instead contrasts tense laryngealized stops with two types of more or less long-lag stops (i.e., lax and aspirated). This study reviews how the mastery pattern has been described in many transcription-based studies of Korean stops, and how it is related to the prediction made by cross-linguistic pattern of the mastery order based on VOT.

1.1 Predicting the mastery order using VOT

Using VOT to describe the adult-like realization of stop phonation categories has made it possible to capture several consistent trends in children's speech acquisition of laryngeal contrasts across languages and to generalize these trends as a well-motivated developmental universal. As noted earlier, one common finding in the mastery of stop consonants across languages is that the voiceless unaspirated stop is mastered before any other contrastive stops Jakobson (1968) stated this mastery order as the implicational universal that "so long as stops in child language are not split according to the behaviour of the glottis, they are generally produced as voiceless and unaspirated" (p.14), which predicts that aspirated or voiced categories are mastered after the voiceless unaspirated category when a language has a contrast involving either aspiration or voicing. As demonstrated in studies of various languages, it is the short lag VOT that typically serves as an undifferentiated acoustic form of children's early stops across languages no matter which laryngeal contrast the language has in its stop consonant (Macken & Barton, 1980; Allen, 1985; Davis, 1995; Pan, 1994).

Again, as noted earlier, Kewley-Port & Preston (1974) relate the earlier mastery of short lag VOT category to the more precise motor control that is required to meet the challenging aerodynamic conditions for voicing lead and long lag VOT. The production of long lag VOT requires a precise temporal adjustment between the oral articulation and the glottal opening gesture. The temporal coordination needs to be precise enough to align the maximum glottis with the oral release. In the same vein, voiced stops (particularly in initial position) are the last category to be acquired and this is predicted by their challenging motoric demands. The production of lead VOT requires that the glottal adduction gesture be made prior to the oral constriction release and also that supra-glottal air pressure be sufficiently lower than the sub-glottal pressure in order for vocal fold vibration to begin during the closure (Westbury, 1983; Westbury & Keating, 1986; Keating, 1983). When one end of the vocal cavity is closed, it is not easy to maintain a supra glottal pressure that is lower than the sub glottal pressure. Compared to the long lag VOT and the lead VOT, the production of short lag VOT is relatively easier in that it is achieved by the glottis opening at any time during the oral occlusion. Moreover, it can even be achieved without opening the glottis, if other maneuvers for relieving intra oral-air pressure are not performed. Before mastering the precise temporal control between the gestures of the glottal opening and the oral constriction release for long lag VOT or lead VOT, children might be making imprecise coordinations between the articulatory gestures producing the acoustic consequence of short lag VOT value.

If Korean children's early stop productions have short lag VOT, as predicted from young children's stop productions across languages, the tense stop will be the adult category that corresponds most closely to the short lag stops produced by children. The VOT of the Korean tense stop is in the short lag region. The three phonation types in Korean have been described as showing at least some VOT overlap in earlier studies (Lisker & Abramson, 1964; Kim, 1965: Han, 1970). These earlier studies documented an overlap between the lax and tense stops, in the short lag VOT range, with the aspirated stops showing values in the long lag VOT range and clearly separated from the other two phonation types. However, recent studies have shown a different pattern of VOT overlap in which the tense stops in the short lag VOT range are clearly separated from lax and aspirated stops at the long lag VOT range. This different pattern of VOT overlap in Korean stops has been described as a diachronic sound change in Silva (2006), Wright (2007) and Kang & Guion (2008). Whereas older Korean speakers have VOT distributions that make the lax stops an intermediate VOT type which overlaps more with the tense stops, Korean speakers born after the 1970s tend to produce lax stops with long lag VOT values that separate them from tense stops but merge them on the VOT dimension with aspirated stops. Therefore, the prediction faithful to the cross-linguistic mastery pattern is that the tense category will be mastered first in Korean children's stop productions.

This prediction accords with the prior transcription-based studies on the mastery of Korean stops phonation types. According to the results of transcription analysis in cross-sectional and longitudinal studies of phonological development in Korean stop transcription analysis, tense, lax and aspirated stops were all mastered before three years, although the tense category seems to be mastered somewhat earlier than the aspirated category. For example, Kim & Pae (2005) conducted norming studies for consonants in various prosodic conditions with Korean-speaking children aged 2;6 to 6;5. They found evidence that tense stops were mastered before lax or aspirated stops, in that word initial /p'/ and /t'/ were produced correctly by 95% of children before 2;6. All three were mastered by 75% of children in the age group from 3;1 to 3;6. Pae (1994) found that the tense and lax stops were in the consonant inventories of children younger than 2;6, and tense stops were commonly substituted for lax stops. Kim (2008) transcribed multiple repetitions of target word-initial stops per child. Tense stops /p', t', k'/ were produced with 75% accuracy by more than 75% of children at 2;6, while the same accuracy for aspirated stops /p^h/, /t^h/ and /k^h/ and lax /t/ were not achieved until 3;0 and 4;0. In a longitudinal case study by Jun (2007), word-initial tense stops were recorded to appear at 17 months, before aspirated (18 mo.) and lax (20 mo.).

1.2 A puzzling pattern of mastering Korean stops

While one could conclude that the mastery pattern of anomalous three-way contrast in Korean stops is not exceptional to the cross-linguistic pattern of mastery, others view the early mastery of Korean tense stop as a puzzling pattern due to the greater motoric demands required to produce tense stops. In this respect, tense stops are unlike voiceless unaspirated stops in other languages, which, as discussed above, are less articulatorily demanding than either voicing lead or long lag stops.

Unlike the production of short lag VOT categories in other languages, the acoustic consequence of short lag VOT in Korean tense stops is the result of a laryngeal muscle tenseness that suppresses the vocal fold vibration despite a closed glottis before the oral constriction release and of a general vocal track tension that restrains the passive expansion of the supra-glottal cavity for lowering the pressure (Kagaya, 1974; Hardcastle, 1973; Hirose, Lee & Ushijima, 1974; Dart, 1987). That is, in the production of tense stops, vocal cord vibration does not begin as soon as complete glottal adduction is met.

Contrastive with the "marked" laryngeal setting of tense stops, the two other types of Korean stops have more cross-linguistically typical laryngeal conditions. Lax stops are characterized by a small glottal opening with a lack of adduction muscle activity. The glottal gestures are not precisely timed to make a mild aspirated stop in initial position, producing the passive voicing of intervocalic lax stops in Korean. The aspirated stop is distinguished from the lax and tense stop categories by having adduction muscle activity, which is suppressed during the closure and becomes active quickly at the release. This results in a glottis held open for a long enough interval to have a period of turbulent aspiration noise. Therefore, it is surprising that Korean-speaking children master the tense category before the lax stop.

Most importantly, the muscle tenseness of tense stop affects not only VOT but also other acoustic dimensions such as fundamental frequency (f0) and voice quality. In addition to having a short lag VOT, the Korean tense stop is further characterized as having a higher f0 immediately after voice onset and a somewhat creaky voice quality, due to the vocal cords being pressed (Kagaya, 1974; Hardcastle, 1973; Hirose, Lee & Ushijima, 1974; Dart, 1987; Jun, 1993). This segmental property of having a higher f0 is phonologized in the intonation structure of Seoul Korean, so that the tonal pattern at the beginning of the accentual phrase is determined by the phonation-type of the phrase-initial consonant segment (Jun, 1993, 1998). The tense or aspirated consonant phonation-type is associated with a H tone-initial accentual phrase (HHLH) while the lax phonation-type is associated with a L tone initial accentual phrase (LHLH). The glottal configurations of the Korean stops affect the following vowel quality in such a way that the vocalic onset after the tense stop has a pressed quality, whereas the vocalic onset after the lax stop has a breathy quality and only the aspirated stop has an unspecified or modal voice quality. When H1-H2 (the amplitude difference between the first and second harmonics; Holmberg, Hillman & Perkell, 1988; Hanson & Chuang, 1999) is measured to capture the voice quality in the vowel just after lax, aspirated, and tense stops, the lax stop is characterized as having the greatest H1-H2 value and the tense stop as the smallest (i.e., negative) H1-H2 value. This acoustic parameter of H1-H2 can distinguish the tense from the lax and the aspirated stops in Korean in adult speakers' productions (Cho, Jun & Ladefoged, 2002; Kim, Beddor & Horrocks, 2002; Kim, 2008). Thus, VOT is only one of several acoustic characteristics that differentiate the tense stops from lax and aspirated stops. Given the multiplicity of acoustic cues that are associated with the "marked" glottal setting for the tense stop, we cannot say that the VOT characteristics of the three different types explains the early mastery of tense stops relative to lax stops unless we can show that native speakers weigh VOT values more than other features in assimilating children's stop productions to the adult norms.

1.3 Hypothesis

One hypothesis that relates the mastery pattern based on the transcription to acoustic characteristics of Korean stops is that the native speaker's transcribed accuracy of children's stops mostly depends on VOT values of the tokens despite the multiplicity of acoustic parameters that differentiate the tense type from the other two types in Korean. This is supported by the changing role of VOT in differentiating the three-way phonation-type contrast found in the adult stop productions in recent studies. If VOT is a sufficient acoustic parameter for cueing the tense category in Korean, and a Korean adult transcriber judged the accuracy of tense stops based on VOT, with little influence from other acoustic parameters such as f0 and H1-H2, a young children's short lag VOT tokens may be heard as tense stops, even though the child is not producing the voice quality that characterizes these stops in adult productions.

This hypothesis can be tested by assessing the relative role of VOT and other acoustic cues in determining the transcription accuracy of tense, lax and aspirated categories in Korean. Therefore, the goal of this study is to examine the relationship between the acoustic properties of Korean children's stop productions and the categories given by a native transcriber (a common assessment of mastery order). We will investigate 1) which acoustic parameter in Korean stops best determines the accuracy of stop phonation-type in children's productions when the stop tokens are judged by a native transcriber and 2) how the acoustic parameters affect adult Korean naive listeners' identification of the phonation-type categories in children's stop productions. Specifically, in the production experiment, the stop productions by about 70 Korean-speaking children (aged 24 through 72 months) and adults (aged 18 through 30 years) were analyzed in terms of the acoustic parameters of f0, H1-H2 and VOT and regression models were used to examine the relationship between the acoustic properties of each token and the transcribed phonation type (for children) or the target phonation type (for adults). A perception experiment followed to examine how the perceptual judgments of naive Korean speaking adults related to the acoustic characteristics.

2.0 Production Experiment

2.1 Method
2.1.1 Materials

The target consonants were word-initial lingual stops in three following vowel contexts (/i/, /a/ and /u/). Three words were selected to elicit the target in each context. For example, Korean /t/ vs. /t'/ vs. /tʰ/ in the /a/ context were elicited in /taŋ.gɨn/ 'carrot', /t'al.gi/ 'strawberry' and /tʰa.dʒo/ 'ostrich'. The presentation order of the words was determined by a randomizing algorithm that insured that the words with the same sequence of target consonant and vowel were not adjacent to another in the list and that at most two words for any CV type were presented in either of two blocks. We chose target words that were familiar to young children (see Table 1 for the lists of word used to elicit the target lingual stops). Multiple tokens of the target words were recorded by an adult female native speaker of Korean using a child directed speaking style. Only tokens that at least four adult listeners repeated correctly were selected as stimuli in the experiment. In addition to the audio stimuli, culturally appropriate pictures of target words were presented along with the audio stimuli in the production experiment.

Table 1 ***A list of Korean words that were elicited in the word repetition task to Korean-speaking children and adults. The sounds were transcribed in IPA.***

IPA	gloss	IPA	gloss
tal.pʰɛŋ.i	snail	ka.baŋ	bag
tan.tʃʰu	button	kam.dʒa	potato
taŋ.gɨn	carrot	kaŋ.adʒi	puppy
ta.ram.dʒwi	squirrel	ka.wi	scissor

tiŋ.gul.diŋ.gul	rolling about	ki.tʃ^{h}a	train
tiŋ.doŋ.dɛŋ	bell-ringing sound	ki.rin	giraffe
tuŋ.gɨl.gɛ	round	ku.du	shoe
tul	two	kuk.dʒa	scoop
tu.bu	tofu	kuk.su	noodle
t’al.raŋ.i	noise maker	k’ak.t’u.gi	radish kimchi
t’aŋ.k^{h}oŋ	nut	k’a.ma.gwi	black bird
t’al.gi	strawberry	k’aŋ.t^{h}oŋ	can
t’a.rɨ.rɨŋ	tinkling	k’a.man.sɛk	black color
t’iŋ.doŋ	door-bell sound	k’i.wʌ.jo	sticking in
t’i.t’i.p’aŋ.p’aŋ	car-honking	k’iŋ.k’iŋ	groaning sound
t’uŋ.bo	plump person	k’ul.k’ʌk	gulping sound
t’uk. t’uk	dripping sound	k’ul.bʌl	honey bee
t’u.kʌŋ	lid	k’ul. k’ul	oink-oink
t^{h}a.ol	tire	k^{h}a.dɨ	card
t^{h}a.i.ʌ	towel	k^{h}al	knife
t^{h}a.dʒo	ostrich	k^{h}a.me.ra	camera
t^{h}ak.dʒa	table	k^{h}a.rɛ	curry
t^{h}i.bi	television	k^{h}i.wi	kiwi
t^{h}i.sjʌ.tʃ^{h}ɨ	T-shirt	k^{h}i.da.ri	tall man
t^{h}u.dʌl.t^{h}u.dʌl	grumbling	k^{h}u. k^{h}i	coolie
t^{h}uk.t^{h}uk	beating around	k^{h}ul. k^{h}ul	snoring
t^{h}uŋ.t^{h}uŋ	stamping	k^{h}u.sjʌn	cushion

2.1.2 Subjects

There were 80 child participants, aged 2;0 to 5;11 (years;months) and 20 adult participants, age 18 to 30 years (see Table 2). Both adults and children were recruited and recorded in Seoul, Korea. All the child and adult participants passed a hearing screening.

Table 2 ***The age and sex distributions of child and adult participants.***

	2; 0 - 2; 11	3; 0 - 3; 11	4; 0 - 4; 11	5; 0 - 5; 11	adults
female	12	11	7	2	10
male	9	9	8	6	11

2.1.3 Task

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

The child subjects were also given an articulation test (Urimal Test of Articulation and Phonology: U-TAP; Kim, 2004), and a receptive vocabulary test (Picture Vocabulary Test; Kim, 1995) in order to make sure their language development was age-appropriate. Recordings were made in childcare centers or in the children’s homes for the youngest children.

2.2 Analysis
2.2.1 Transcribed accuracy/error analysis

If children produced more than one token, only the first production was included in the analysis. All of the productions were transcribed by the first author, a trained phonetician and a native speaker of Korean. First, the transcriber coded the consonant as correct or incorrect. Then, the transcriber provided an alphabetic transcription of those productions that were judged as incorrect. While non-plosive productions were regarded as errors, place of articulation errors were coded as correct, as long as the phonation-type categories were correct. For example, [t] or [tʃ] for a target /k/ was counted as correct for voicing despite the error in the place of articulation. We excluded errors of deletion, distortion and non-plosive productions from the analysis. 10% of the data were transcribed by the second native-speaker transcriber in order to assess the inter-transcriber reliability. Phoneme-by-phoneme inter-rater reliability was above 90%.

The mastery pattern of each stop phonation-type category was predicted in two apparent-time analyses based on the transcribed accuracy and error pattern using mixed effects logistic regression models (Snijder & Bosker, 1999; Roudenbush & Bryk, 2002). The model of accuracy analysis predicts the log-odds of the transcribed accuracy of each token based on children's age in months and consonant phonation types. Equation 1 and Equation 2 show the formula of the two models. The first model predicts the log-odds of the transcribed accuracy (*Accuracy*) of each token produced by a speaker, based on children's age in months (*Age*) and phonation types (*Phon.Type*). The random intercept and slope of the phonation types in the speaker level were added to the model, denoted as U_0 and U_1 in Equation 1.

Equation 1

$$\log\left(\frac{Accuracy}{1 - Accuracy}\right) = \gamma_{00} + \gamma_{10} Phon.Type + \gamma_{20} Age + U_0 + U_1 Phon.Type$$

In order to further test whether the tense phonation type was most common in younger children's productions, the model of error analysis was made that predicts the log odds of probability of being substituted as a tense category given the lax and the aspirated targets by age. The log odds of being substituted with tense category given the target of lax or aspirated stops were predicted by the children's age in months (*Age*) and the target consonant phonation types (lax and aspirated stops: *TargetCat*). The model included the random intercept and slope of the target consonant phonation types in the speaker level.

Equation 2

$$\log\left(\frac{Subst.Tense}{1 - Subst.Tense}\right) = \gamma_{00} + \gamma_{10} TargetCat + \gamma_{20} Age + U_0 + U_1 TargetCat$$

2.2.2 Acoustic analysis

VOT was measured by subtracting the time of burst (i.e., the beginning of an abrupt energy rise after the closure) from the time of the first indication of voicing, evident from the voicing bar in the spectrogram, as well as a visible initiation of the first regular cycle of periodicity in the waveform.

The f0 was measured by taking the reciprocal of the interval between two neighbouring pulses at 20 ms after the voicing onset. The glottal pulses were automatically detected using the pulse function in Praat (this is an auto-correlation based periodicity detector). The f0 measurement was made at 20 ms after the voicing onset instead of immediately at the voicing onset because it was observed that the function was more reliable (higher correlation coefficient) several glottal pulses after the exact onset of voicing.

We used H1-H2 as our spectral tilt measurement to capture the breathiness at the vowel onset. It was measured by subtracting the amplitude (in dB) of the second harmonic from the amplitude of the first harmonic in the Fast Fourier Transform spectrum generated based on a 25 ms analysis Hamming window beginning at the voicing onset. The frequency location of the first harmonic was automatically detected by Praat, referring to the fundamental frequency as an initial value, and then the researcher checked the precise frequency locations for the first and the second harmonic in a token-by-token manner in order not to be confused between DC noise and H1 or between H2 and A1 (amplitude of the first formant). There were instances where there was no clear second harmonic due to weak periodicity at the voicing onset. For these tokens, the frequency at twice the first harmonic was taken to be the frequency of the second harmonic, and the amplitude at that frequency location was taken as H2.

A mixed effects logistic regression model was used to evaluate the effects of each acoustic parameter in predicting the consonant categories as expressed in Equation 3. To accommodate the three-way contrast of Korean stop phonation, we had two equivalent sets of mixed effects logistic regression models. One set of mixed effects logistic regression models predicted the tense vs. the non-tense types (i.e., the lax and the aspirated types) as the dependent variable, and the other set of models predicted the lax vs. the aspirated types as the dependent variable. In each set of model, the log odds of the probability of being transcribed as one particular phonation type (e.g., tense versus non-tense: *Phon.Type*) were predicted by three acoustic parameters (*VOT*, *f0* and *H1H2*) that characterize each token and by the variance of these acoustic parameters that exist in the individual speaker level as the random effects. We assumed that the adults accurately produced the target stops without phonation-type errors, and therefore, we treat the target phonation types of adult productions as equivalent to the transcribed categories of the child productions.

Equation 3

$$\log\left(\frac{Phon.Type}{1-phon.Type}\right)=\gamma_{00}+\gamma_{10}VOT+\gamma_{20}f0+\gamma_{30}H1H2+U_0+U_1VOT+U_2f0+U_3H1H2$$

The coefficients (i.e., γ_{10}, γ_{20}, γ_{30} ...) of each parameter indicate the size of the effect in determining the tenseness of the item in the tense versus non-tense model, if the coefficient is a significantly effective variable in the model. The greater the absolute value of the coefficient is, the more influential is the variable. To remove the magnitude differences among measurement units (millisecond, decibel and hertz), the three acoustic parameters were standardized using z-score transformation. The model considers that there are random effects of intercept and slope coefficients at the individual speaker level, denoted as '*U*' in Equation 3.

The effects of the three acoustic parameters were estimated in three separate models of different speaker groups (children, adult females and adult males). This was done to control for sex and age differences (e.g., adult males have lower f0 and generally less breathy voice quality than females and children due to the morphological changes in the larynx at puberty).

2.3 Results

2.3.1 Transcription accuracy/error analysis

The left panel in Figure 1 presents the pattern of transcribed accuracy in Korean. Recall that we analyzed only the children's productions that were transcribed as plosives (i.e., excluding deletions, distortions and substitutions of sounds other than non-plosives). In this figure, the percentage of each individual child's target stops that met these criteria that were judged to have the target phonation type is plotted as a function of the child's age in months. When the inverse logit function curves (generated based on the output of the mixed effects logistic regression model) are overlaid on top of the scatter-plot, the overall trend is clear. While all three phonation type categories are produced accurately at a relatively early age, tense stops are produced more accurately than the other two types at the younger ages. The regression curves from the analysis predict that tense stops will be produced with 77% accuracy at 24 months, while lax and aspirated stops will be produced with 69% and 70% accuracy, respectively at the same age.

FIGURE 1

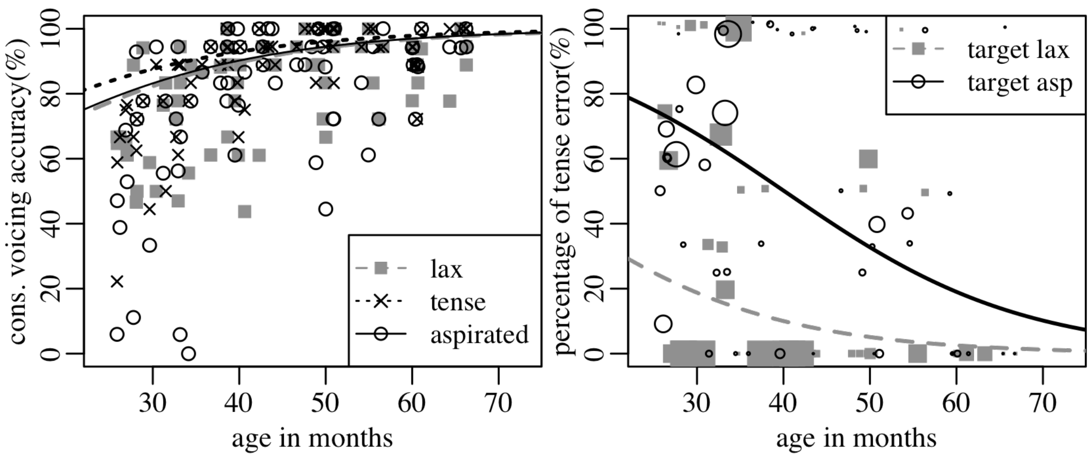


FIGURE 1 caption:

Transcribed accuracy of phonation-type categories for children's word-initial stops in Korean (left figure) and a scatter-plot of percentages of substituted tense categories for errored productions of non-tense categories (i.e., [tense] for /lax/ and /aspirated/) as a function of the child's age in months (right figure). In the right figure, the size of the plotting character indicates the number of relevant errors that the child made in the production. The curves are the inverse logit curves predicted by the mixed effect logistic regression models (Table 3 and Table 4).

According to the output of the model, as summarized in Table 3, the effect of lax type of consonant ('lax') was significantly different from the effect of tense stops (a reference category). The effect of aspirated type of consonant ('aspirated') in predicting the production accuracy was significantly different from that of tense stops, although the significance was only marginal at $p < 0.1$. The significant main effect of age suggests that the transcribed accuracy of each type of consonants increases as age increase.

Table 3 ***The output of mixed effects logistic regression model that predicts the production accuracy of phonation types spoken by Korean-speaking children: Equation 1 and Figure 1 (left panel). The tense stop was the reference category of the model.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr	
subject	(Intercept)	0.52	0.72		
	aspirated	1.40	1.18	0.22	
	lax	1.10	1.05	-0.63	-0.30

Number of observations: 3410, groups: subject, 67

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	2.48	0.14	17.48	< 0.001 ***
age	0.06	0.01	7.24	< 0.001 ***
aspirated	-0.35	0.21	-1.69	0.09 .
lax	-0.41	0.19	-2.16	< 0.05 *

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 4 ***The output of mixed effects logistic regression model that predicts the substituted category of tense stops for the intended targets of aspirated or lax stops: Equation 2 and Figure 1 (right panel). The age was centered at the grand mean of children's age in months, which is 38.5 months.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
subject	(Intercept)	2.35	1.53	
	lax	8.96	2.99	-0.41

Number of observations: 381, groups: subject, 63

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
age	|-0.07	0.02	-3.14	< 0.01 **
aspirated	0.12	0.29	0.40	0.68
lax	-2.08	0.51	-4.06	< 0.001 ***

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The analysis of transcription accuracy found that the Korean tense stops were only slightly more accurate than lax and aspirated stops. The accuracy differences across the three phonation categories were relatively smaller than those found in earlier transcription studies of Korean stops. For example, in Kim (2008), only 50% of children aged 2;6-3;0 produced at least 75% of /t/ target tokens accurately, whereas all children of the same age group produced 75% of /t/ targets correctly. The relatively small accuracy difference between tense stops and lax or aspirated stops can be attributed to the judgment criteria of the accuracy where only phonation-type errors were counted as incorrect in the transcription.

The substitution patterns in the Korean children's error productions in the current data also support the idea that the tense stop appears before the aspirated stop, because a tense stop was commonly substituted for a target aspirated stop in the younger children's productions. The right panel of Figure 1 shows the percentage of errors on a child-by-child basis where a tense stop was substituted for a lax or aspirated stop (i.e., the number of tense substitutions divided by the number of total errors) as a function of the child's age in months. The curves show that there was a greater likelihood for a tense type to be substituted for a target aspirated stop at younger ages and the use of the tense stop as the substitution category decreased over time. Children were more likely to substitute an aspirated stop for the lax target in errored productions indicated by the curves laid below 50% of tense substitutions at y-axis. However, there was a trend toward fewer such substitutions at younger ages (i.e., more tense substitutions for the youngest children).

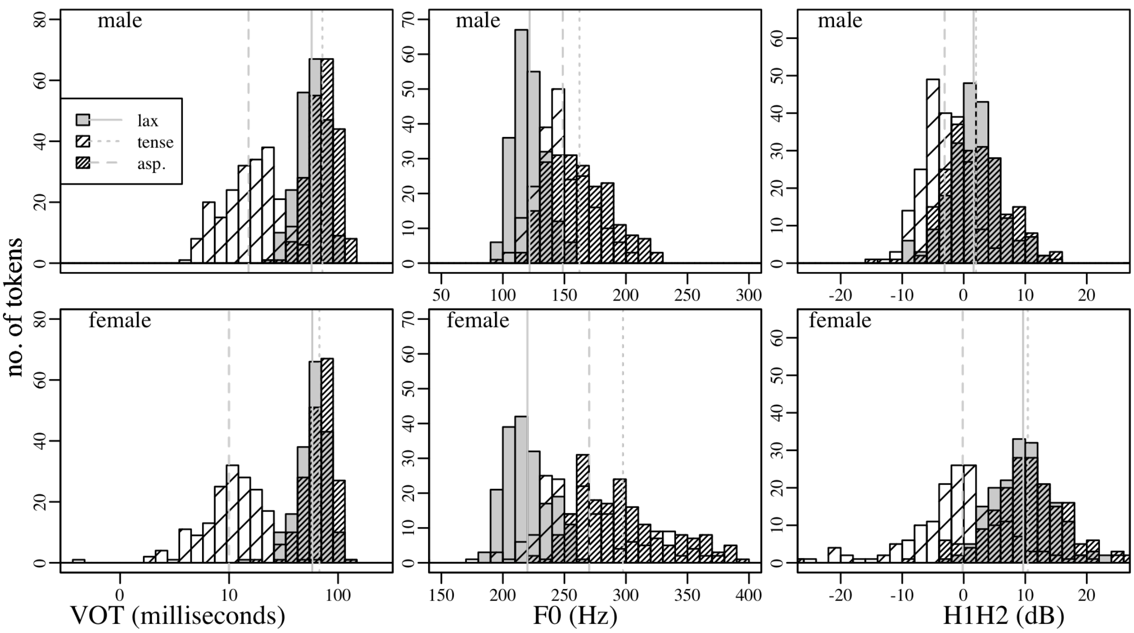
2.3.2 Acoustics characteristics: adult productions

Figure 2 shows the values of VOT, f0 and H1-H2 separated by gender (female vs. male) for productions of adult speakers. The VOT histograms in the leftmost panels show that the tense stops occupied the short lag VOT range in the adults' stop productions. While the VOT range for the tense stops overlapped minimally with the ranges for the lax or aspirated stops, lax stops shared a wide range of VOT values with aspirated stops. This almost complete overlap in VOT values between lax and aspirated stops with no overlap between lax and tense stops seen in studies from the 1960s is consistent with the results of Silva (2006), Wright (2007) and Kang & Guion (2008), where the longer lax stop VOTs were identified as indicating a sound change in progress. Our adult speakers were all younger than 30 years and so they all fall in the category of younger Korean speakers born since the 1970s who tend to have longer VOT values in the lax stops.

It was along the f0 dimension that lax stops were differentiated from the aspirated stops by having lower values. As shown in the histograms in the middle panels in Figure 2, f0 values at vowel onset after lax stops were lower than those after aspirated stops.

The right-hand panels of Figure 2 show histograms for H1-H2, the breathiness measure. Although H1-H2 values for all three types produced by Korean female speakers were greater than those produced by male speakers, tense stops have smaller H1-H2 values than lax and aspirated stops for the productions of both female and male speakers.

FIGURE 2



FIFUGRE 2 caption:

Histograms of VOT, f0 and H1-H2 for lax, tense and aspirated stops produced by Korean adult male and female speakers. Vertical lines indicate median values.

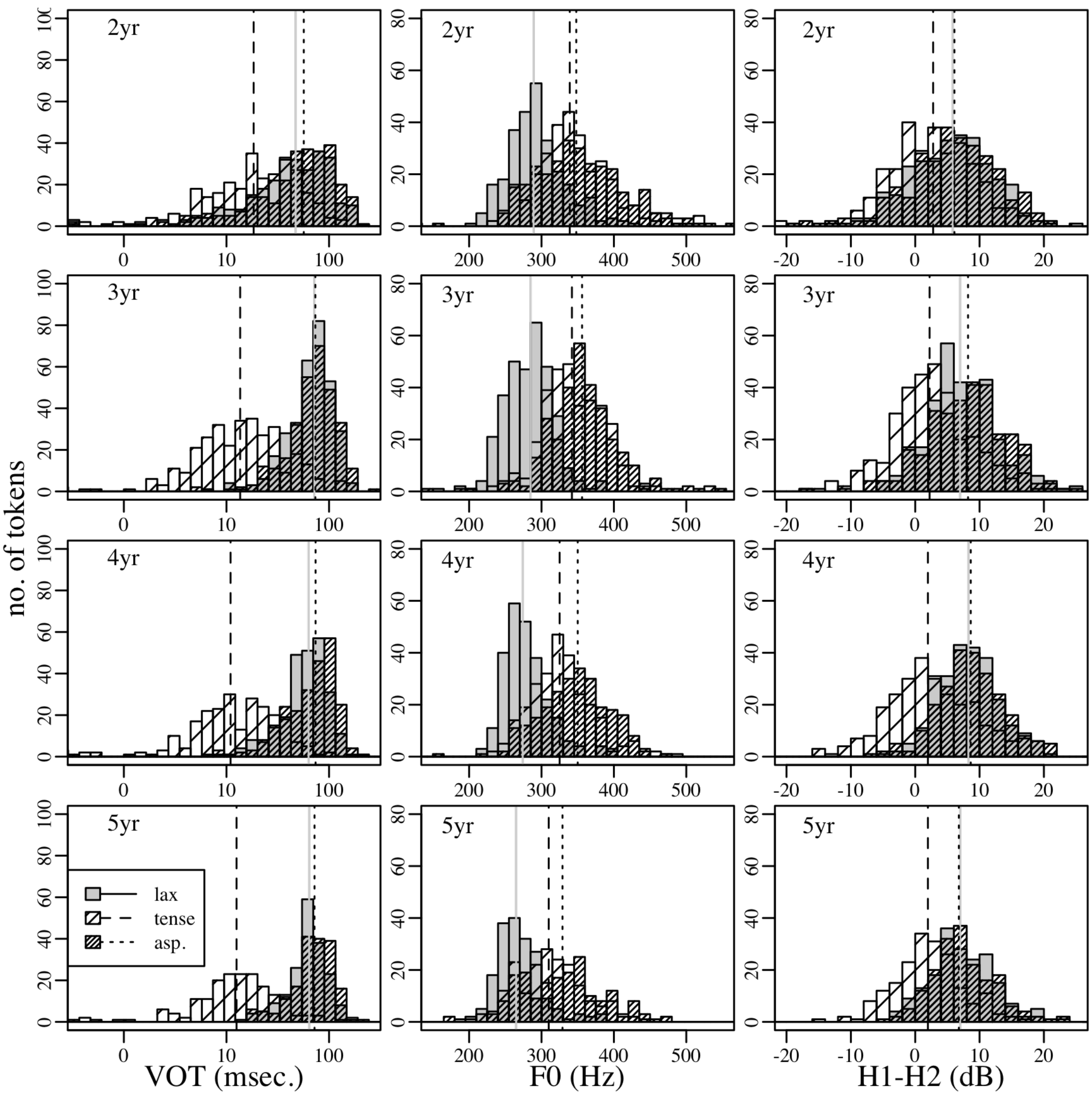
2.3.3 Acoustic characteristics: child productions

Figure 3 shows the distributions of VOT, H1-H2 and f0 of Korean-speaking children's stop productions. In the youngest age group (2;0-2;11) shown in the top leftmost panels, the VOT values of tense stops were concentrated at a short lag range with relatively sharp peaks skewed toward zero VOT. By contrast, the lax and aspirated stops had relatively more variability in VOT values, covering both the short and the long lag ranges. In the older children's productions (3;0-5;11), the lax and aspirated stops had distributional peaks at longer VOT values that were clearly separated from the peaks for tense stops. At all ages, the medians for the lax stops were not greatly different from those for the aspirated stops, which is similar to the pattern in Korean adult speakers' VOT distributions. Among the three different phonation types, the VOT values of the tense stops were most adult-like in the 2-year-olds' productions in that they were realized as short lag VOT values despite wider variability. However, unlike adults' tense stops, there were a small number of tense stops that were made with lead VOT values in Korean-speaking children's productions. The duration of the prevoicing lead in children's tense stops was relatively short compared to the lead VOT for voiced stops of other languages studied in Lisker and Abramson (1964), which ranged from -170 ms to -45 ms in Spanish, for instance.

The distributions of f0 values for tense, lax and aspirated stops (middle panels) show that, while there was some overlap among the three types of stops along the f0 range, the lax stops were mostly distributed at a lower f0 range and the tense and aspirated stops were distributed at a higher f0 range where they almost completely overlapped with each other. This was true even for 2-year-olds' stop productions.

In the 2-year-olds' H1-H2 distributions (top right panel), there was no distinction among the three types of stops. Values for all three types almost completely overlapped with each other. In the older children's productions, the H1-H2 values showed some separation between the tense stops and the other two types. The H1-H2 values of tense stops were generally lower than those of lax and aspirated stops. In addition, although the median H1-H2 values of all three types of stops were positive, tense stops had the most tokens with negative H1-H2 tokens.

FIGURE 3



FIFUGRE 3 caption:

Histograms of VOT, f0 and H1-H2 in stops produced by Korean-speaking children by age group. Vertical lines indicate median values.

2.3.3 Acoustic characteristics: Mixed effects logistic regression models

As expressed in Equation 3, the mixed effects logistic regression models were constructed to examine how these acoustic characteristics are integrated into the transcriber's judgment of phonation types. Recall that we have two sets of model. In one set, we contrasted tense stops to the other two types (i.e., lax and aspirated stops) and in the other set, we contrasted lax stops to aspirated stops.

Tense vs. Non-tense: The output of the mixed effects logistic regression models is summarized in Table 5 for children, Table 6 for adult males and Table 7 for adult females, and the associated figures are shown in the upper panels of Figure 4

Table 5 ***The output of the mixed effects logistic regression model that predicted the transcribed tense category in contrast with the non-tense types (i.e., the lax and aspirated categories) in the productions of children.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
subject	(Intercept)	0.49	0.70			
	VOT	2.91	1.71	-0.09		
	f0	0.16	0.41	0.10	-0.03	
	H1-H2	0.18	0.43	-0.25	-0.09	0.03

Number of observations: 3110, groups: subject, 67

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-0.53	0.11	-4.71	< 0.001 ***
VOT	-3.75	0.26	-14.43	< 0.001 ***
f0	0.60	0.09	6.56	< 0.001 ***
H1-H2	-0.60	0.09	-6.57	< 0.001 ***

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 6 ***The output of the mixed effects logistic regression model that predicted the transcribed tense category in contrast with the non-tense types (i.e., the lax and aspirated categories) in the productions of adult males.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
subject	(Intercept)	5.53	2.35			
	VOT	0.08	0.27	1.00		
	f0	0.08	0.29	-1.00	-1.00	
	H1-H2	0.78	0.88	-0.18	-0.18	0.18

Number of observations: 529, groups: subject, 10

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-2.17	0.90	-2.40	< 0.05 *
VOT	-10.77	1.84	-5.83	< 0.001 ***
f0	0.80	0.45	1.75	0.08 .
H1-H2	-1.37	0.58	-2.35	< 0.05 *

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 7 ***The output of the mixed effects logistic regression model that predicted the transcribed tense category in contrast with the non-tense types (i.e., the lax and aspirated categories) in the productions of adult females.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
subject	(Intercept)	0.90	0.95			
	VOT	11.39	3.37	-0.08		
	f0	1.91	1.38	0.59	0.75	
	H1-H2	0.80	0.89	-0.99	0.22	-0.47

Number of observations: 530, groups: subject, 10

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-2.16	0.63	-3.39	< 0.001 ***
VOT	-10.85	2.17	-4.99	< 0.001 ***
f0	-0.79	0.81	-0.97	0.32
H1-H2	-0.50	0.55	-0.91	0.36

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

As shown in Table 5 there were significant main effects of coefficients associated with VOT, f0 and H1-H2 in the model of children's productions. The absolute value of the coefficient associated with VOT was greatest among the three acoustic parameters, suggesting that the discrimination of tense from non-tense was more sensitive to VOT than to f0 or H1-H2. The negative signs of the coefficients in Table 5 simply indicate that the prediction of tense stops is negatively correlated with the acoustic values of VOT and H1-H2. That is, the higher values of VOT and H1-H2 resulted in a higher probability of the transcriber identifying the tokens as one of the non-tense types.

Unlike the model of children's production, the two models of adult male and female speakers showed that not all three acoustic parameters were significant in differentiating the tense category from the non-tense categories. It was only VOT that has a significant effect in the adult female productions as shown in Table 7 while it was VOT and H1-H2 that have significant effects in the adult male productions as shown in Table 6. Despite these differences, it was consistent that the models estimated a greater coefficient of VOT than those of f0 and H1-H2 across all three speaker groups. Because the values of the three acoustic parameters were standardized, the larger VOT coefficient relative to the H1-H2 coefficient serves as evidence that VOT plays the dominant role in the differentiation of tense from non-tense stops.

The absolute values of the coefficients determine the steepness of the slopes of the curves in Figure 4. The steepest slopes of VOT in the models for all three groups in the leftmost panels of Figure 4 show how effective VOT was in differentiating between tense and non-tense stops for all three participant groups.

FIGURE 4

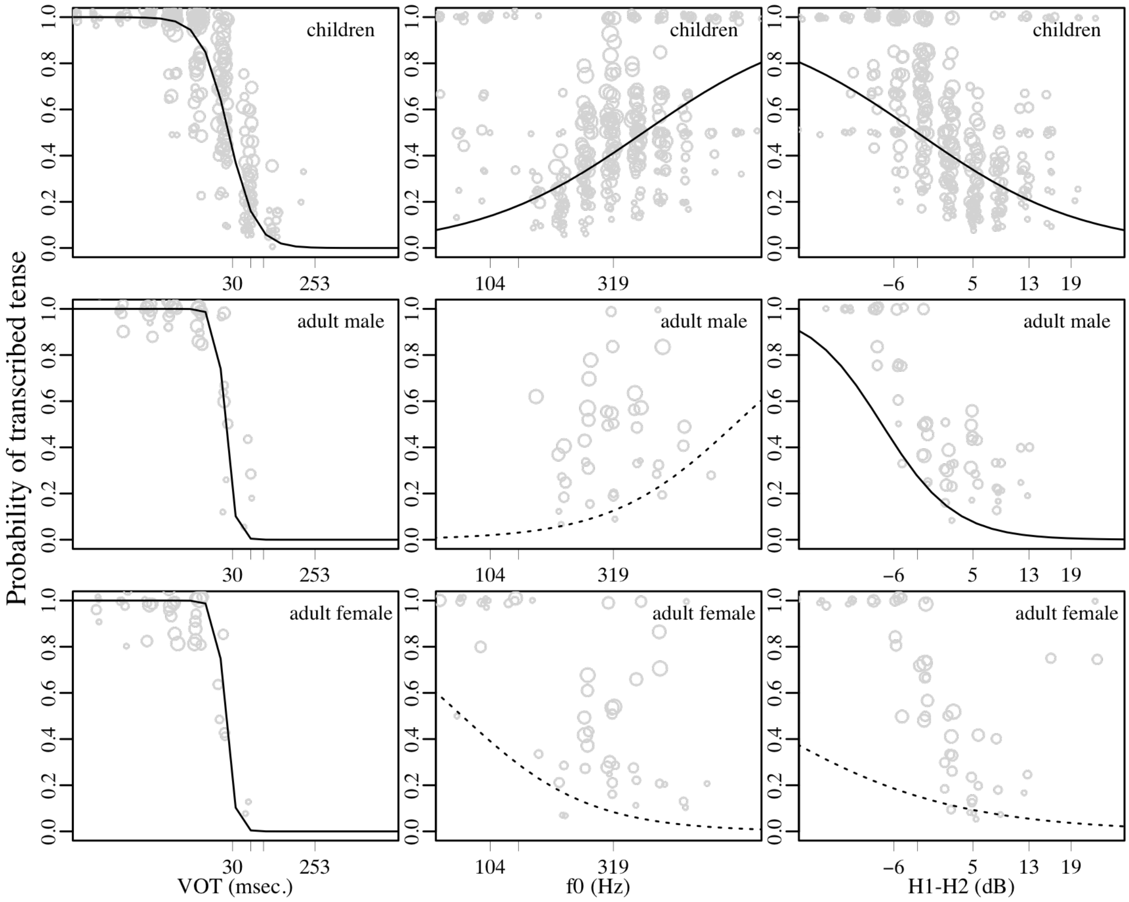


FIGURE 4 caption:
The scatter-plots of the probability of tense stops with respect to VOT, f0 and H1-H2 parameters estimated by the mixed effects models of logistic regression. The plots are separated by the three speaker groups (i.e., child, adult male and adult female). Each data point of the plots represents a proportion of individual speaker's tense type tokens to the stop productions, whose acoustic values fall in a particular bin along the specified acoustic dimension. The curves are the inverse logit curves drawn based on the output of the mixed effects logistic regressions shown in Table 5, Table 6 and Table 7. The solid curves in the figures indicate that a particular acoustic parameter was a significantly effective predictor of the contrast between tense and non-tense stops according to the model. The dashed curves indicate non-significant predictors.

Lax vs. Aspirated stops: The output of the logistic regression model is presented at Table 8 for children, Table 9 for adult males and Table 10 for adult females and the associated figures are shown in Figure 5.

Table 8 ***The output of the mixed effects logistic regression model that predicted the transcribed lax type in contrast with the aspirated type in stop productions of children.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
subject	(Intercept)	1.58	1.26			
	VOT	1.57	1.25	-0.79		
	f0	2.81	1.67	0.01	0.009	
	H1-H2	0.15	0.39	0.002	0.47	0.37

Number of observations: 2001, groups: subject, 67

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	1.17	0.20	5.80	< 0.001 ***
VOT	-2.51	0.25	-10.03	< 0.001 ***
f0	-2.64	0.24	-10.94	< 0.001 ***
H1-H2	-0.12	0.09	-1.31	0.19

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 9 ***The output of the mixed effects logistic regression model that predicted the transcribed lax type in contrast with the aspirated type in stop productions of male adults.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
subject	(Intercept)	23.41	4.83			
	VOT	3.63	1.90	-0.32		
	f0	7.53	2.74	0.28	-0.99	
	H1-H2	0.50	0.71	-0.15	-0.88	0.90

Number of observations: 355, groups: subject, 10

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	1.13	1.75	0.64	0.51
VOT	-4.43	1.42	-3.10	< 0.01 **
f0	-8.12	1.38	-5.88	< 0.001 ***
H1-H2	-0.60	0.46	-1.30	0.19

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 10 ***The output of the mixed effects logistic regression model that predicted the transcribed lax type in contrast with the aspirated type in stop productions of adult females.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
subject	(Intercept)	1.52	1.23			
	VOT	1.01	1.00	0.28		
	f0	2.57	1.60	-0.55	-0.41	
	H1-H2	0.04	0.20	0.44	0.36	-0.03

Number of observations: 360, groups: subject, 10

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	2.18	0.74	2.93	< 0.001 ***
VOT	-3.19	0.99	-3.21	< 0.001 ***
f0	-5.44	0.83	-6.51	< 0.001 ***
H1-H2	-0.30	0.44	-0.69	0.48

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

As shown in Table 8, the coefficients of the fixed effects in the model of children's productions suggest that there were significant main effects of VOT and f0 but no significant effects of H1-H2 in differentiating the lax stops from the aspirated ones produced by children. The results were consistent in the other two adult production models, showing that the effects of VOT and f0 were significant but not that of H1-H2 was significant as presented in Table 9 and Table 10. In all three models of lax vs. aspirated, the absolute values of coefficients of f0 were higher than coefficients of VOT. This suggests that the f0 parameter was more influential than the VOT parameter in differentiating the lax stops from the aspirated stops across the three speaker groups. It is noted that, among the three groups, the absolute values of VOT coefficient was most close to that of f0 in the child model. The absolute value of coefficients of f0 was only slightly higher than coefficient of VOT.

The absolute values of coefficients determine the steepness of the slopes in Figure 5, showing that the bottom two panels for adult male and female speakers' productions have the curves with steeper slopes of the f0 parameter than the VOT parameter. In the child speaker model (top panels), the slope of the VOT parameter was as almost steep as that of the f0 parameter. The identification of lax stops was negatively correlated with the two parameters of f0 and VOT, as indicated by the signs of the coefficients. The increase of f0 and VOT values resulted in higher probability of the transcriber identifying the tokens as the aspirated type.

FIGURE 5

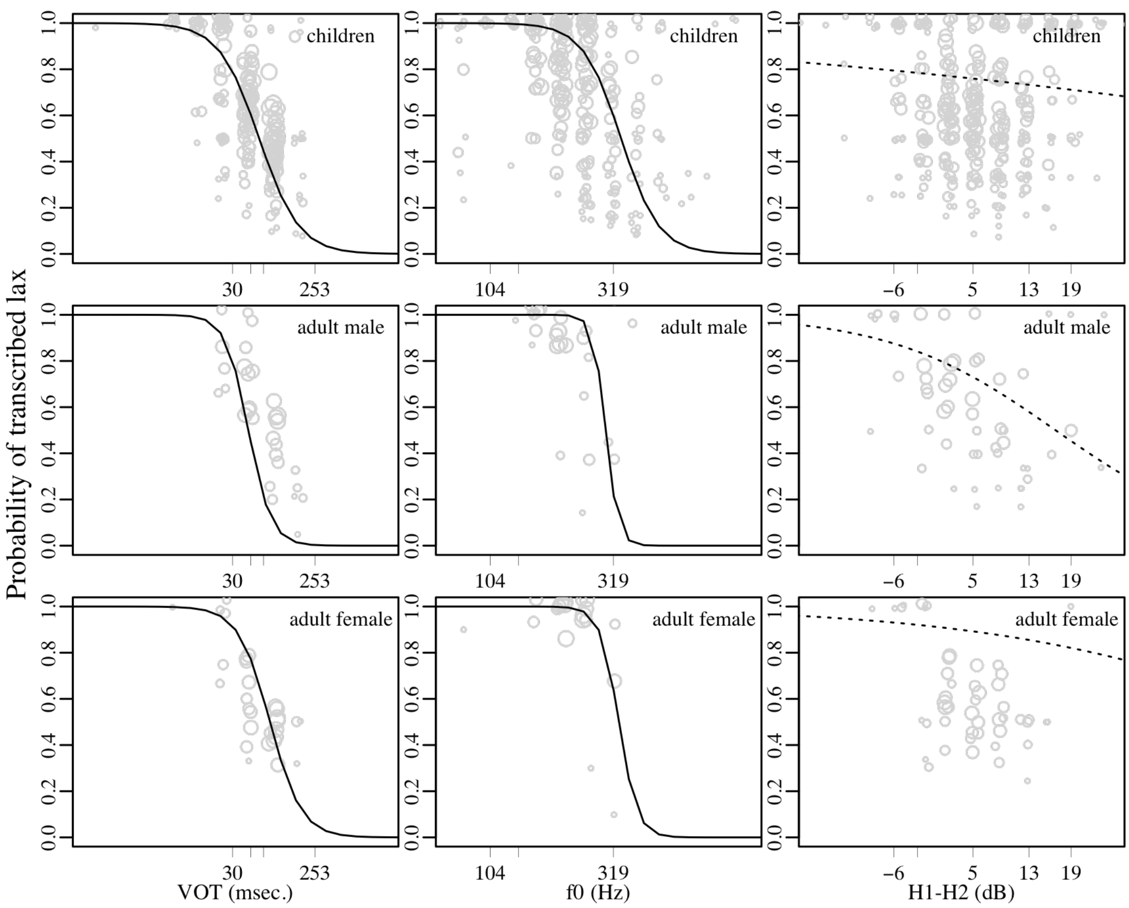


FIGURE 5 caption:

The scatter-plots of the probability of lax stops with respect to VOT, f0 and H1-H2 parameters estimated by the mixed effects models of logistic regression. The plots were separated by the three speaker groups (i.e., child, adult male and adult female) in each column. The curves are the inverse logit curves drawn based on the output of the mixed effects logistic regressions shown in Table 8, Table 9 and Table 10. The solid curves in the figures indicate that a particular acoustic parameter was a significantly effective predictor of the contrast between lax and aspirated stops according to the model. The dashed curves indicate non-significant predictors.

2.4 Discussion

We examined the transcribed accuracy of the three different Korean stops of the laryngeal contrast and the acoustic characteristics of the stops produced by adults and children with specific goals in mind. First, we wanted to examine the time-course of acquisition of the three phonation-categories in the productions of 2- to 5-year-old children. Second, we examined how the three phonation categories were differentiated acoustically in adults' productions. Third, we were interested in investigating how the acoustic characteristics of children's stop productions influenced the transcriber's judgment of production accuracy. Given that the single acoustic dimension of VOT cannot discriminate among all three phonation types in Korean stops, we examined two acoustic dimensions in addition to VOT in order to investigate how the acoustic characteristics of children's stop productions influenced the judgment of transcribed categories.

Consistent with the findings of prior transcription studies, the early mastery of tense stops in children's productions was observed in the current cross-sectional production study. The early mastery of tense stops was observed in its higher transcription accuracy and in the youngest children's use of the tense type as a substitution category.

The adult patterns of acoustic characteristics in the three different types of stop productions suggested that VOT differentiates tense stops from non-tense stops, but it does not distinguish lax from aspirated stops. The distinction between the lax and the aspirated stops was instead made by a combination of VOT and f0. The lax stops have a lower f0 than the aspirated stops. These acoustic characteristics reflect a recent diachronic change in the role of VOT in the distinction between the three types of stops as suggested by Silva (2006), Wright (2007) and Kang & Guion (2008).

The patterns of acoustic parameters in children's stop productions resembled the adult pattern of contrasting tense, lax and aspirated stops, in that younger children tended to produce stops with short lag values of VOT and lax stops were characterized as the lower f0 values. As in other languages, all three categories of Korean stops had short lag VOT values in the youngest children's productions and the lax and aspirated stops were differentiated from the tense stops in older ages by having longer VOT values.

The new production pattern of stop phonation-type contrast affected the native transcriber's judgments of the accuracy of the tense versus non-tense categories as they were produced by children. A stop with the short lag VOT was transcribed as correct for the tense stop more readily than for the non-tense stops. The transcribed categories for children's lax and aspirated stops tended to be "tense" for short lag values, and otherwise largely determined by the f0 value, differentiating the productions that were transcribed as lax stops from those that were transcribed as aspirated stops.

According to the mixed effects logistic regression model, the differentiation of transcribed tense categories vs. non-tense ones were predicted more effectively by VOT than f0 or H1-H2. The set of mixed effects logistic regression models for differentiating lax and aspirated stops showed that the variations of transcribed lax vs. aspirated stops were explained by the f0 and VOT parameters. While the effect of f0 was greater than VOT in this model, VOT was almost as effective as f0 in differentiating lax and aspirated stops in the children's production model.

The early mastery of tense stops as determined by native-speaker transcription can be explained by the role of VOT in the transcription judgments. As in previous research, the youngest Korean-speaking children in this study tended to produce stops with short-lag VOT values. In addition, VOT was the primary acoustic cue that the native-speaker/transcriber used to differentiate tense from non-tense in children's stop productions. Due to this primary role of VOT in differentiating tense from non-tense stops, the early mastery of tense stops, a short-lag VOT category in Korean, is not such an exceptional pattern after all.

A shortcoming of the first experiment is that the categorization of the children's productions into the three phonation-type categories was made by a single native speaker who was a trained phonetician and who could listen to each stimulus item multiple times and also examine the waveform and spectrogram. In perception experiment, we examine the role of the same three acoustic cues in differentiating the three phonation-type categories for stop consonants using a larger group of Korean adults who are naive listeners and who had only a single opportunity to listen to each token. The categorization task was planned to make the perception experiment comparable to a trained phonetician's transcription task. The category goodness judgement task was included in the perception experiment in order to examine whether the acoustic parameters influence the listeners in defining a good exemplar of the selected category.

3.0 Perception Experiment

3.1 Method
3.1.1 Materials

A subset of adults and children's tokens of /t/- /t^h/- and /t'/-beginning words used in the production study was chosen to be presented as stimuli to 20 Korean-speaking adult listeners. Only the consonant and vowel portion was excised from the selected words to avoid any perceptual bias from the lexicon. The following vowel context varied among /a/, /i/ and /u/. The stimuli include tokens produced by children and adults. Table 11 shows the distributions of talker's age and vowel context of the stimuli.

Table 11 ***The age and vowel distributions of 400 stimuli.***

	t			t'			t^h			
	/a/	/i/	/u/	/a/	/i/	/u/	/a/	/i/	/u/	total
2; 0 - 2; 11	22	1	6	13	2	7	20	3	12	85
3; 0 - 3; 11	19	4	10	17	6	10	20	3	7	96
4; 0 - 4; 11	9	12	8	16	6	14	20	4	8	97
5; 0 - 5; 11	13	5	10	8	6	6	17	4	2	71
adults	7	5	5	8	3	6	6	4	6	50
total	70	27	5	8	3	6	6	4	6	400

In order to balance the number of tokens which were likely to be perceived as /t/ or /t^h/ or /t'/ in Korean, the single native speaker transcriber's judgments of the token were referred to as a rough estimation of plausible perceptual outcomes. The 400 stimuli (350 from children's production and 50 from adults' production) were chosen primarily based on the stop VOT values, aiming to sample the VOT distribution so as to reflect the whole range of the natural data obtained from the production experiment. Figure 6 shows the distributions of VOT, f0 and H1-H2 in all of the stimuli used in the experiment. The quantiles of stimulus set of child speakers' tokens were compared with those of the children's production data in each acoustic parameter as shown the bottom panels in Figure 6. The distributions of f0 and H1-H2 also turned out to reflect the range of the natural data after sampling the stimuli based on the VOT variation. Another 10 practice items were chosen, with VOT covering the same range, to be presented at the beginning of the experiment as practice trials.

FIGURE 6

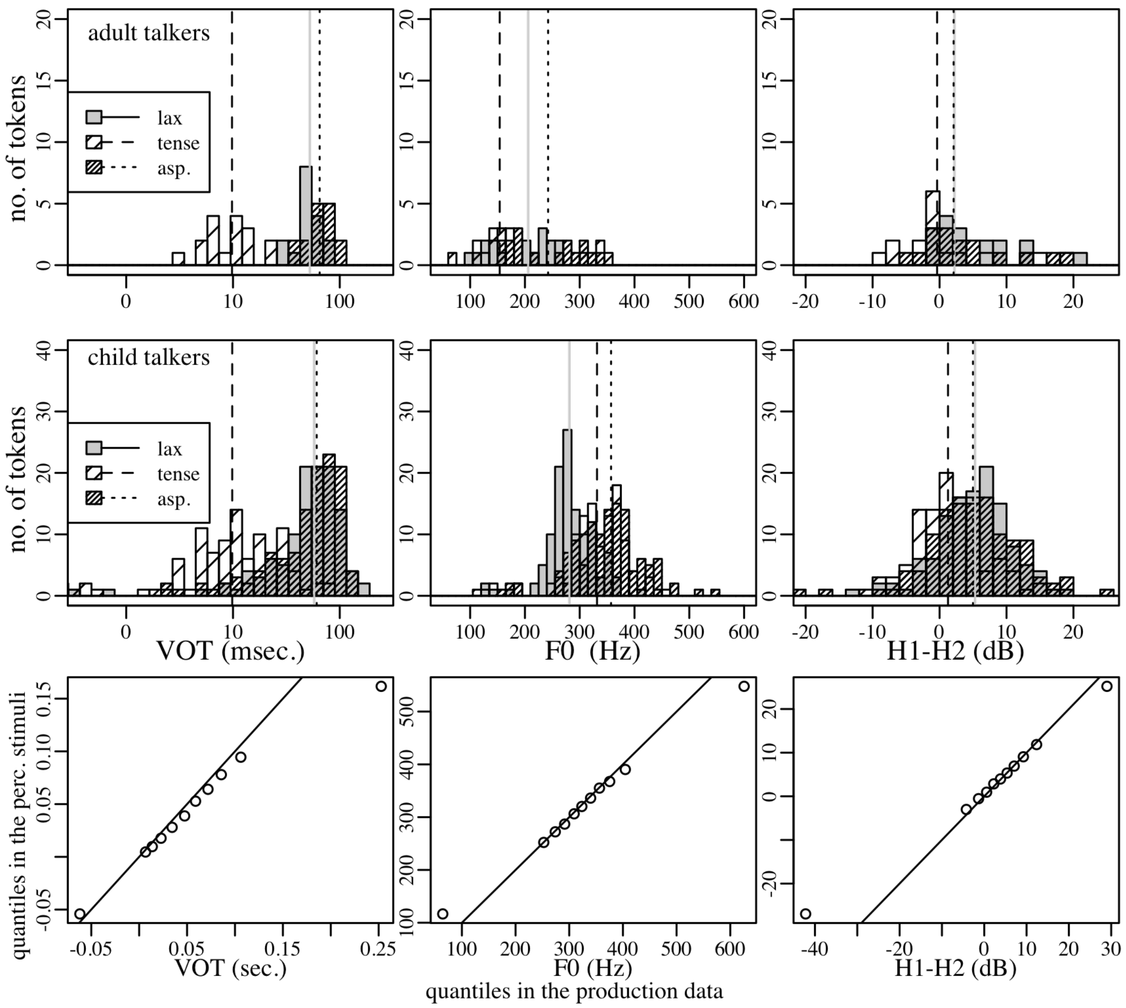


FIGURE 6 caption:
The histograms of VOT, f0 and H1-H2 of the 400 stimuli used in the perception experiment. 50 of the total stimuli were from the adult speakers' word productions (top panels) and 350 of them came from the child speakers' word productions (middle panels). The bottom panels are the quantiles of perception stimuli distributions of child talkers and children's production data used in Section 2 against each other. The pattern that quantile values are aligned on diagonal lines suggests that the distributions of the three acoustic parameters are similar between two sets of data.

3.1.2 Subjects and Task

Twenty college students at the Kyunghee University in Seoul participated in this experiment. They were all Seoul Korean speakers with no reported history of speech, language or hearing disorders.

After each stimulus item was played, the Hangul orthographic symbols for /t/, /t'/ and /t^h/ appeared on the computer monitor. Subjects were asked to select one stop category as their choice by clicking the letter on the screen using a mouse. Immediately after the response to the stop categories, a double-ended arrow was presented on the computer monitor, and subjects were asked to click on a location along the arrow to indicate the gradient goodness of each stimulus. The two end points of the arrow were labelled as 'very good exemplar' and 'very poor exemplar' in Hangul. An optional break was given after every 100 stimuli. Listeners were paid for their participation.

3.1.3 Analysis

Comparable sets of mixed effects logistic regression models were made to describe the relationship between the perceptual categorical responses to the adults' and children's stop productions and the acoustic parameters VOT, f0 and H1-H2 associated with the stimuli. The models predict the stop phonation type perceived by adult listeners to the stimuli produced by adults and children. Similar to the logistic regression models in the production experiment, the acoustic parameters of VOT, f0 and H1-H2 were the fixed effects and the 'listener' factor was specified as a random effect allowing random intercepts and slopes. This is expressed in Equation 4.

Equation 4

$$\log\left(\frac{Phon.Type_{perc.}}{1-phon.Type_{perc}}\right)=\gamma_{00}+\gamma_{10}VOT+\gamma_{20}f0+\gamma_{30}H1H2+U_0+U_1VOT+U_2f0+U_3H1H2$$

The effects of these acoustic parameters were estimated in the three different models of talker groups (child-, adult male- and adult female-talkers of the stimuli). As done in Section 2.0, we made two sets of such models to predict the tense vs. non-tense stops (i.e., lax and aspirated stops) in one model, and the lax vs. aspirated stops in the other model. The three acoustic parameters were standardized using z-score transformations in order to remove the magnitude differences among measurement units (millisecond, decibel and hertz). The z-score normalization was done across all the child and adult talker stimuli.

The goodness ratings of each stimulus were examined as a function of each acoustic parameter in the individual phonation category identified by listeners. The linear regression analysis was performed to predict the listeners' ratings based on each of VOT, f0 and H1-H2. The dependent variable was listeners' judgments encoded as the pixel number clicked along the arrow in the monitor. VOT, f0 and H1-H2 were independent variables in the linear regression model.

3.2 Results
3.2.1 Categorical judgements

Tense vs. Non-tense stops: Table 12, Table 13 and Table 14 and Figure 7 show the output of the mixed effects logistic regression models, and the associated figures of the results. According to the fixed effect coefficients in the model of the responses to child taker stimuli, all three acoustic parameters of VOT, f0 and H1-H2 were significantly effective predictors in differentiating perceptual judgments of tense stops from those of non-tense stops. The primary predictive role of VOT was evidenced by the greatest absolute value of the VOT coefficients among the three parameters, suggesting that the listeners' judgments of tense stops were determined primarily by VOT. In the model of child talker stimuli perception, the coefficient for VOT (4.3) was almost 14 times greater than the coefficient for f0 (0.3) and 8 times greater than H1-H2 (0.55). While the VOT and H1-H2 dimensions were negatively correlated with the perceptual judgment of tense, the f0 was positively correlated with tense perception. The primary role of VOT was illustrated in Figure 7 by the fact that the VOT parameter curve has the steepest slope.

In the models of two other stimulus talker groups (adult males and adult females), VOT was the primary acoustic parameter in differentiating the judgements of tense stops from those of non-tense stops. While either f0 or H1-H2 was not a significant acoustic parameter in judgements of tense stops vs. non-tense stops in the stimulus adult male/female talker models, as shown in Table 12 and Table 13, VOT remained as the significant acoustic parameter in these models. The absolute regression coefficients of VOT was the greatest in the models, which is shown as the steepest slopes of the VOT parameter curves in the bottom two panels of Figure 7.

Table 12 ***The output of the mixed effects logistic regression model that predicts the perceived tense stops in contrast with the non-tense stops (Tense vs. Non-tense stops) produced by children.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
listener	(Intercept)	0.11	0.34			
	VOT	0.60	0.77	-0.50		
	f0	0.0006	0.02	0.19	0.75	
	H1-H2	0.02	0.14	-0.01	0.87	0.97

Number of observations: 7000, groups: listener, 20

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-0.25	0.08	-2.94	< 0.01 **
VOT	-4.29	0.20	-21.14	< 0.001 ***
f0	0.30	0.03	7.65	< 0.001 ***
H1-H2	-0.55	0.05	-10.72	< 0.001 ***

Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Table 13 ***The output of the mixed effects logistic regression model that predicts the perceived tense stops in contrast with the non-tense stops (Tense vs. Non-tense stops) produced by adult male speakers. The model did not allow the random slopes of independent variables in the listener level due to little variability in that level. The estimated log-likelihood of this model was not significantly different from that of the model with the random slopes.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
listener	(Intercept)	< 0.001	< 0.001	

Number of observations: 500, groups: listener, 20

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-1.21	0.32	-3.75	< 0.001 ***
VOT	-4.92	0.61	-8.01	< 0.001 ***
f0	0.23	0.29	0.79	0.42
H1-H2	-1.76	0.63	-2.80	< 0.01 **

Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

Table 14 ***The output of the mixed effects logistic regression model that predicts the perceived tense stops in contrast with the non-tense stops (Tense vs. Non-tense stops) produced by adult female speakers. The model did not allow the random slopes of independent variables in the listener level due to little variability in that level. The estimated log-likelihood of this model was not significantly different from that of the model with the random slopes.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
listener	(Intercept)	< 0.001	< 0.001	

Number of observations: 500, groups: listener, 20

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-2.03	0.44	-4.58	< 0.001 ***
VOT	-6.90	1.09	-6.30	< 0.001 ***
f0	-0.87	0.33	-2.63	< 0.01 ***
H1-H2	0.28	0.42	0.68	0.49

Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

FIGURE 7

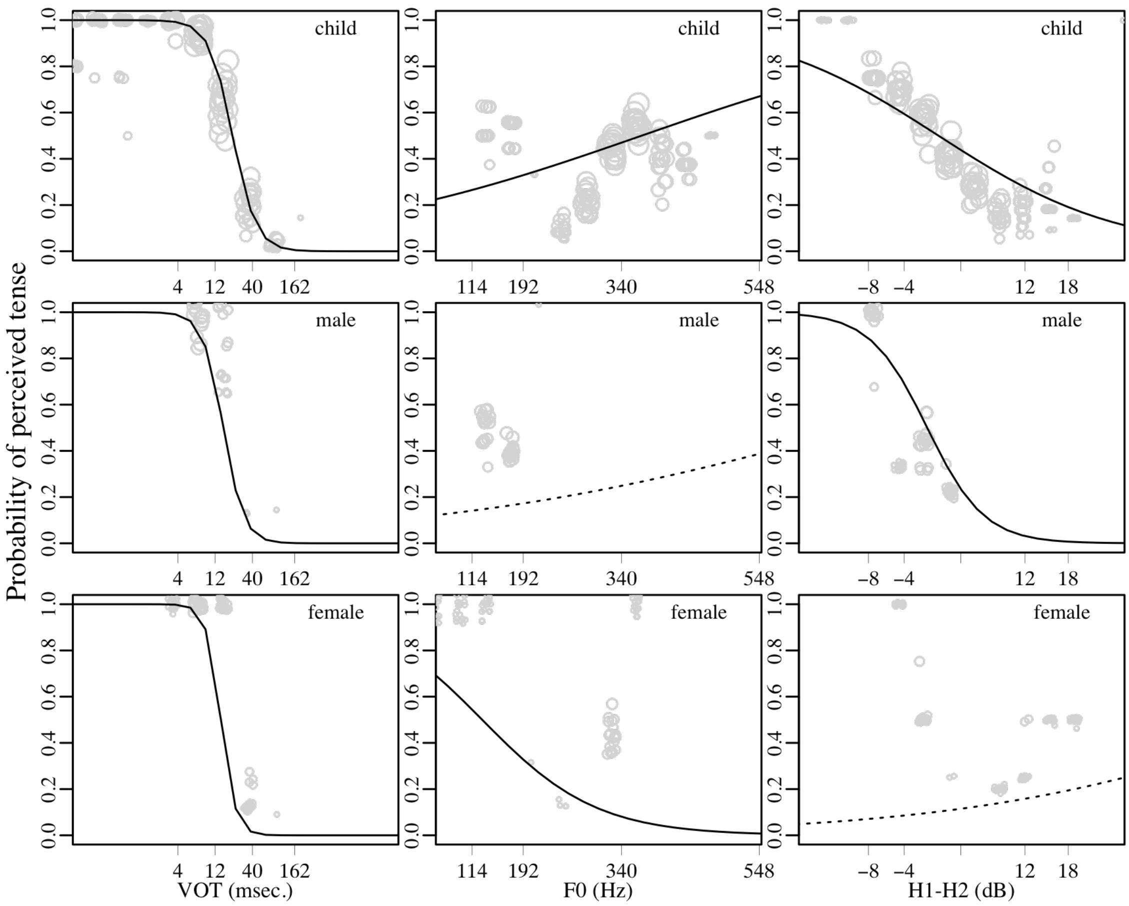


FIGURE 7 caption:
The scatter-plots of the probability of being perceived as a tense stop as a function of its acoustic value of VOT, f0 and H1-H2. The probability is estimated by calculating the number of tense responses divided by the total number of stimuli provided at each bin along each acoustic continuum. The curves are the inverse logit curves drawn based on the output of the mixed effects logistic regressions shown in Table 12, Table 13 and Table 14. The solid curves in the figures indicate that a particular acoustic parameter was a significantly effective predictor of the contrast between perceived tense and non-tense stops according to the model. The dashed curves indicate non-significant predictors.

Lax vs. Aspirated stops: The output of the mixed effects logistic regression models in each stimulus talker group is summarized in Table 15, Table 16 and Table 17 and the associated figures are presented in Figure 8. In the child stimulus talker model as shown in Table 15, VOT and f0 (but not H1-H2) were significant predictors in differentiating perceived lax stops from aspirated stops. The absolute coefficient value of VOT was greater than that of f0, suggesting that the listeners judged the perception of the lax vs. aspirated stops relying more on VOT than f0. According to the signs of the coefficients, the increases of VOT and f0 negatively affected the listeners' judgment of lax stops.

The output of the adult male talker model shows that VOT, f0 and H1-H2 were significant predictors in differentiating perceived lax stops from aspirated stops, as summarized in Table 16. Similar to the child stimulus talker model, VOT has a greater absolute coefficient value than f0 and H1-H2. However, different results were observed for the adult female talker model (Table 17). The absolute coefficient value of f0 was greater than that of VOT in the adult female talker model as, while the effect of VOT parameter failed to reach significance. This is interpreted as indicating that listeners were not sensitive to changes in VOT, but were mainly sensitive to the change in f0 when they identified lax vs. aspirated stops produced by adult female speakers. It is noteworthy that this greater effect of f0 was achieved despite having chosen the stimuli to emphasize variation in VOT.

While the interactions of acoustic parameters between the adult- and child-stimulus types are interesting, the most important finding relative to the questions addressed in this paper is that, even for the contrast between lax and aspirated stops, VOT is the primary acoustic parameter that the listeners used to differentiate these two phonation types when they listened to children's productions. This is illustrated in Figure 8, which shows that the slopes of the VOT parameter curve are steeper than those of the curves for f0 in the child stimulus talker condition (top panels). The slope of the f0 curve is steepest in the adult female stimulus talker condition (bottom panels).

Table 15 ***Output of the mixed effects logistic regression model that predicted perceived lax stops in contrast with aspirated stops (Lax vs. Aspirated stops) produced by children.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr		
listener	(Intercept)	2.03	1.42			
	VOT	0.87	0.93	-0.42		
	f0	0.01	0.10	-0.79	-0.21	
	H1-H2	0.05	0.22	0.71	0.30	-0.97

Number of observations: 4429, groups: listener, 20

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-0.22	0.32	-0.69	0.48
VOT	-2.69	0.25	-10.63	< 0.001 ***
f0	-0.92	0.05	-16.23	< 0.001 ***
H1-H2	0.12	0.07	1.66	0.09 .

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 16 ***The output of the mixed effects logistic regression model that predicts the perceived lax stop in contrast with the aspirated stops (Lax vs. Aspirated stops) produced by adult male speakers.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
listener	(Intercept)	0.98	0.99	

Number of observations: 305, groups: listener, 20

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-0.11	0.41	-0.28	0.77
VOT	-3.31	0.52	-6.36	< 0.001 ***
f0	-0.93	0.19	-4.66	< 0.001 ***
H1-H2	1.20	0.25	4.67	< 0.001 ***

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 17 ***The output of the mixed effects logistic regression model that predicts the perceived lax stop in contrast with the aspirated stops (Lax vs. Aspirated stops) produced by adult female speakers.***

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
listener	(Intercept)	2.56	1.60	

Number of observations: 355, groups: listener, 20

Fixed effects:

	Estimate	Std.Error	z-value	Pr (> |z|)
(Intercept)	-0.86	0.51	-1.66	0.09 .
VOT	-0.17	0.57	-0.30	0.75
f0	-3.96	0.49	-8.08	< 0.001 ***
H1-H2	0.59	0.18	3.26	< 0.01 **

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

FIGURE 8

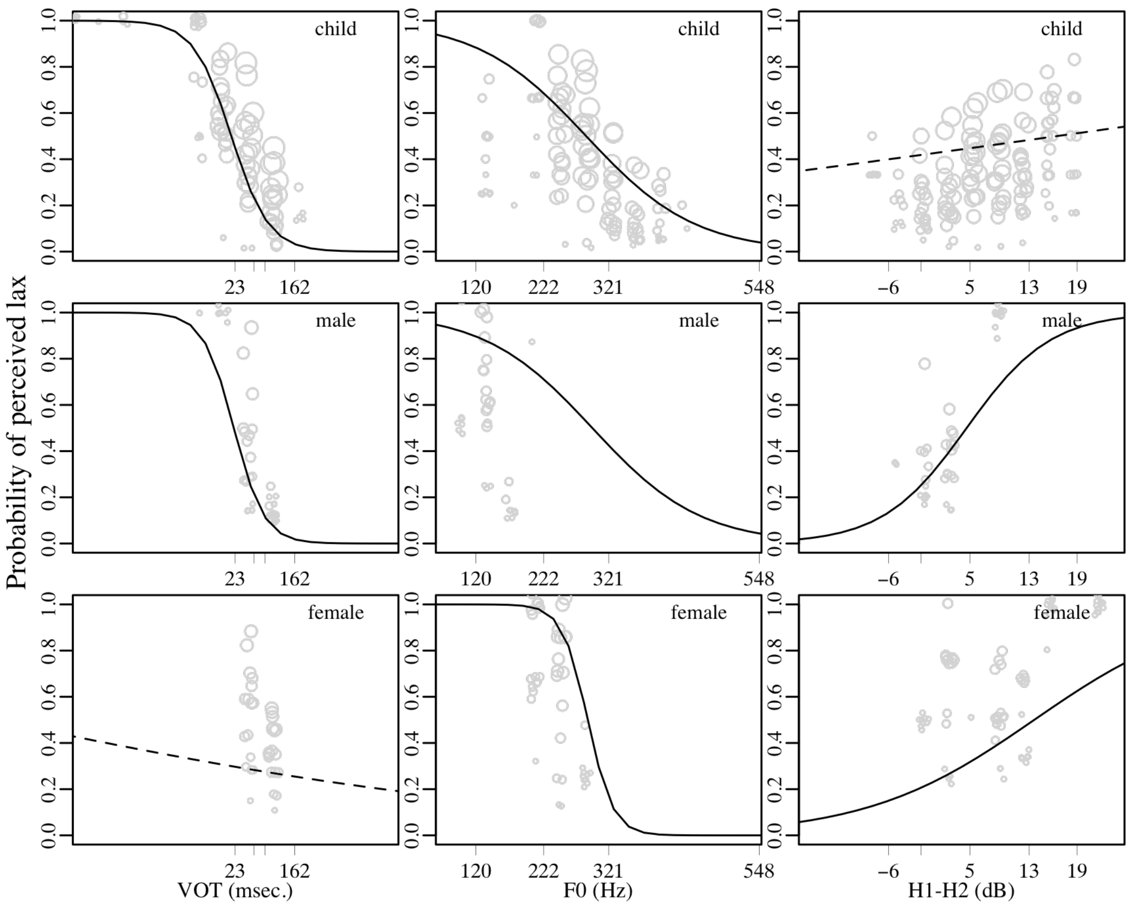


FIGURE 8 caption:

The scatter-plot of probability of perceived lax stops with respect to VOT, f0 and H1-H2 parameters estimated in Korean adults' perception to adults and children's productions. Each data point in the scatter-plots represents a proportion of individual listener's perceived lax stops over the responses whose acoustic values fall in each bin along the acoustic dimensions. The curves are the inverse logit curves drawn based on the output of the mixed effects logistic regressions shown in Table 15, Table 16 and Table 17. The solid curves in the figures indicate that a particular acoustic parameter was a significantly effective predictor of the contrast between perceived lax and aspirated stops according to the model. The dashed curves indicate non-significant predictors.

3.2.2 Within-category goodness judgments

Figure 9 shows the box-plots of the goodness ratings for each of the three talker groups and for each of the three phonation categories. The medians of goodness ratings across the plots were located slightly toward to the 'good' side, which can be explained by the fact that the stimuli were already good enough to be categorized as that particular sound.

FIGURE 9

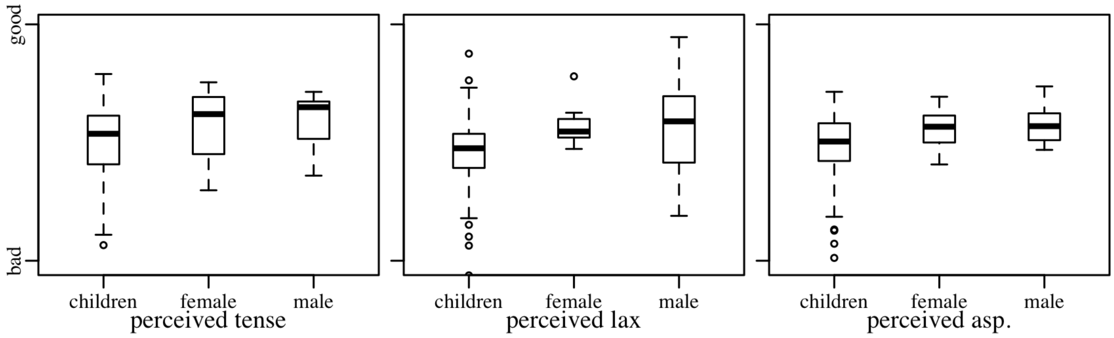


FIGURE 9 caption:

Box-plots of the goodness ratings of stimuli identified as tense, lax and aspirated types, separated by the talker groups (children, adult females and adult males).

The goodness ratings for the each stimulus were averaged across the listeners in order to describe the relationship between the acoustic parameters and the ratings. The VOT parameter was a significant predictor of the variability of the goodness ratings for the tense (R^2= .19) and the aspirated stops (R^2= .33) but not for the lax stops (R^2 < .01). As shown in Figure 10, the tense stops (leftmost panels) tended to be judged as better exemplars as VOT values decreased. In contrast, the aspirated stops (rightmost panels), tended to be judged as better exemplars as VOT values increased. The lax stops are shown in the middle panels, and the lack of a significant relationship between VOT values and goodness ratings can be observed.

H1-H2 also accounted for a small but significant amount of variability in the goodness ratings for the tense stops (R^2= .14), but was not a significant predictor for the lax and the aspirated categories. f0 was not a significant predictor of the goodness ratings for any of the three stop categories.

FIGURE 10

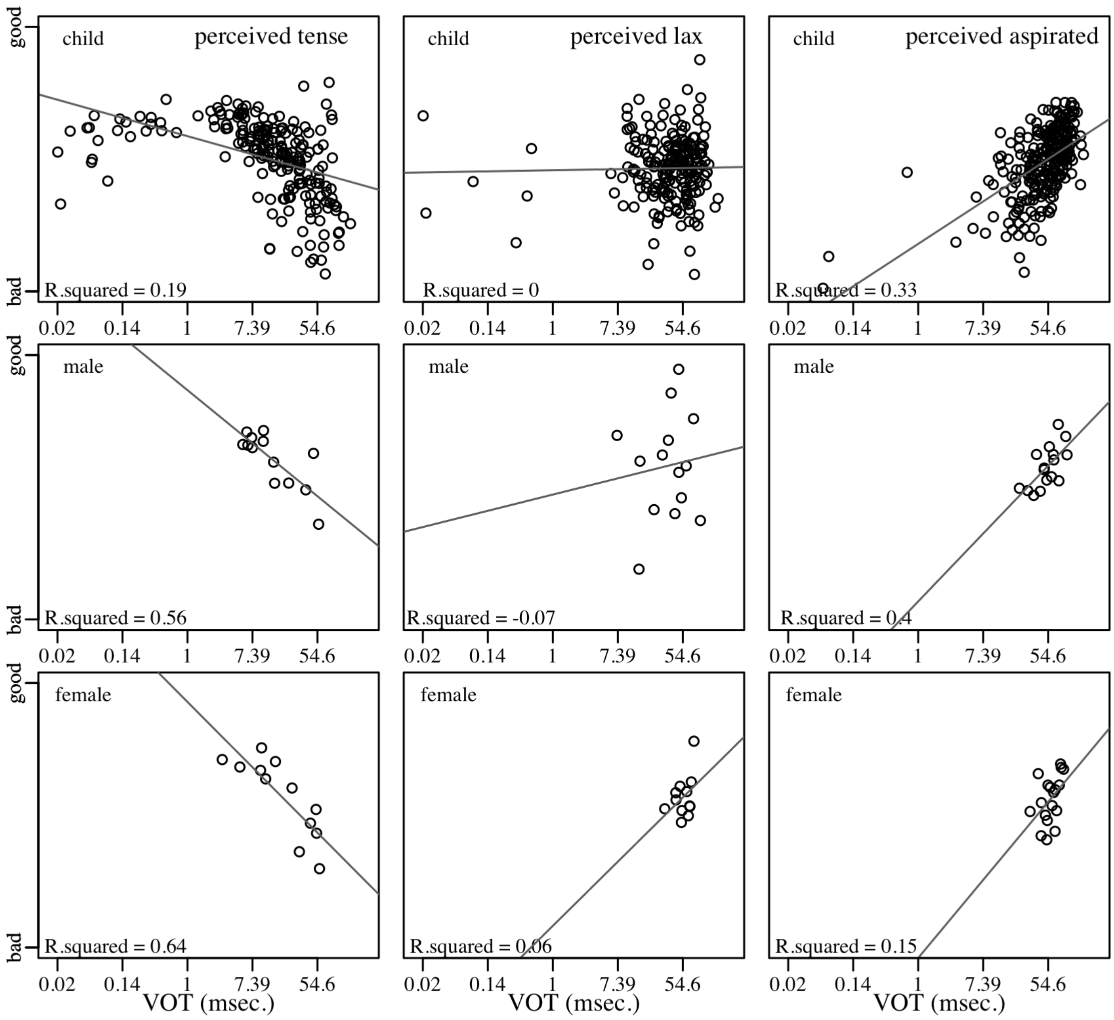


FIGURE 10 caption:

The scatter-plots of the goodness ratings of the category identified as the tense, lax and aspirated types as a function of VOT. The plots were separated by the stimulus talker groups (children, adult males and adult females).

3.3 Discussion

The perception experiment examined naive native-speaking Korean adults' categorization of the three phonation categories of Korean stop consonants. We were interested in the relationship between Korean adults' perception of these three phonation categories and the acoustic dimensions of VOT, f0 and H1-H2 in productions of stop consonants by children and adult males and females. We were also interested in whether the relationship between perception and the three acoustic parameters would be similar for the single trained phonetician whose judgments were analyzed in production experiment and in the 20 naive listeners of perception experiment.

Based on the production experiment, we predicted that the naive listeners would also rely primarily on VOT to differentiate the three phonation types in children's productions. The 20 naive listeners, like the trained transcriber, relied primarily on VOT to differentiate tense from non-tense stops. This was true, regardless of whether the stimuli were produced by children, adult females, or adult males. The other two acoustic parameters, f0 and H1-H2, were also significant predictors, but they had smaller effect sizes.

The results were also similar for the trained phonetician and the 20 naive listeners, when we examined the relationship between perceptual judgments and the three acoustic parameters used to differentiate between lax and aspirated stops. For the children's productions, VOT was the most important cue in categorizing children's stop as lax or aspirated, and f0 was also an important predictor. The relationship between the naive listeners' perception of children's stops and the associated acoustic parameters supports the earlier claim that the early mastery of Korean tense stop can be explained by the way that Korean-speaking children's short lag VOT categories are understood by Korean adults. In both experiments, we found that short lag VOT tokens produced by children were perceived as tense by Korean adults, even though these productions were not necessarily produced with adult-like f0 or H1-H2 values.

VOT influenced listeners' within-category judgments of the tokens identified as tense and aspirated stops. Tokens with VOT values near the category boundaries were judged as poor exemplars by listeners for both tense and aspirated stops. The direction of the relationship was parallel to the categorical judgment results, in that goodness ratings increased as VOT values decreased for tense stops and as VOT values increased for aspirated stops. While VOT was a significant predictor for between- and within-category judgments for tense and aspirated stops, it was not a significant predictor of listeners' goodness judgments for the lax stop category. f0 was also not a significant predictor of goodness ratings for lax stops, although it was the acoustic parameter that best differentiated the lax category from the tense and the aspirated categories in the adult productions. More research is needed to understand why this is the case.

4.0 Conclusion

This paper investigated how Korean children's early mastery of tense stops in transcription-based studies can be related to the role of acoustic properties in categorizing the children's stops. In one sense, the early mastery of Korean tense stop is an exception to the cross-linguistically predicted order of mastery. That is, the conventional wisdom is that whenever a language contrasts a less common consonant type with a more common one, the more common "default" type should be mastered first. This would predict that the tense (or mildly ejective) stops of Korean should be mastered later than the cross-linguistically more common "plain" (or lax) types. The production of tense stops in Korean requires the precise articulatory specifications of posturing the larynx and other articulatory systems for positive vocal cord tension and vocal tract tension, while the lax stop does not require these highly precise articulatory specifications. Therefore, the early mastery of tense stops in the transcription studies poses a puzzle.

However, tense stops in Korean are produced with short lag VOT, which is the VOT category which is generally produced earliest across languages. Because VOT overlaps between lax and aspirated stops (in younger adults) or between tense and lax stops (in older adults), VOT alone cannot differentiate among the three phonation categories in Korean. The two other important acoustic characteristics of tense stops are pressed voice quality and fundamental frequency. We predicted that VOT would play a primary role in differentiating tense and non-tense stops in children's productions and that adult judgments would be less affected by the other acoustic parameters, H1-H2 and f0. This prediction was partly based on the changing characteristics of the Korean stop phonation categories with respect to VOT and f0.

We tested this hypothesis by examining the relationship between the three acoustic parameters and productions of the three phonation types produced by children and adults, and categorized by a trained transcriber and a number of naive adult listeners. We found that the early mastery of tense stops in Korean-speaking children's productions was related to the fact that 1) young Korean-speaking children tend to produce short-lag VOTs in their earliest stop consonant productions (as is common cross-linguistically) and 2) adult Koreans (both naive listeners and a trained phonetician) tend to identify productions with short-lag VOTs as the tense type. Thus, the early mastery of tense stops in Korean is not a cross-linguistically exception when we consider how children's productions are understood by adult listeners of the native language.

Acknowledgements
We thank the children who participated in the study, the parents who gave their consent, and the child-care centers at which the data were collected. This research was supported by NIDCD Grant 02932 to Jan Edwards and NSF Grant BCS0739206 to Mary E. Beckman.

Reference List

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.

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

Clumeck, 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.

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

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.

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.

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

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

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

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

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

Jakobson, R. (1968). *Child Language Aphasia and Phonological Universals*. The Hague:Mouton.

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

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

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

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

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

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

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

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

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

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

Kim, M., 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, Y.T., Chang, H.S., Lim, S.S. & Paek, H.J. (1995). Geurim eohuiryeok keomsa (picture vocabulary test). Seoul: Changaein chonghap pokjikwan.

Kim, Y.T. & Shin, M.J. (2004). *Urimal Test of Articulation and Phonology*. Seoul:Hakjisa.

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

Macken, M.A. & Barton, D. (1980). The Acquisition of the Voicing Contrast in English: A Study of Voice Onset Time in Word-Initial Stop Consonants. *Journal of Child Language*, 7, 41-74.

Pae, S. (1994) Development of normal speech sounds (I): children aged 1;4-3;11, *Understanding and treatment of speech and language disorders: children's articulation disorders*. Seoul: Kunja. 27-53

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

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

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

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

Westbury, J. & Keating, P. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, 22, 145-166.

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. thesis, University of Pennsylvania, Philadelphia, PA.