CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Korean-speaking children's perceptual development in multidimensional acoustic space

Gayeon SON*

University of Pennsylvania, Department of Linguistics, Philadelphia, PA, USA, and Kwangwoon University, Department of English Language and Literature, Seoul, Korea
*Corresponding author: Department of English Language and Literature, 20 Kwangwoon-ro, Nowon-gu, Seoul 01897, Korea. E-mail: gson@kw.ac.kr

Abstract

This study investigated how Korean toddlers' perception of stop categories develops in the acoustic dimensions of VOT and F0. To examine the developmental trajectory of VOT and F0 in toddlers' perceptual space, a perceptual identification test with natural and synthesized sound stimuli was conducted with 58 Korean monolingual children (aged 2–4 years). The results revealed that toddlers' perceptual mapping functions on VOT mainly in the high-pitch environment, resulting in more successful perceptual accuracy in fortis or aspirated stops than in lenis stops. F0 development is correlated with the perceptual distinction of lenis from aspirated stops, but no consistent categorical perception for F0 was found before four years of age. The findings suggest that multi-parametric control in perceptual development guides an acquisition ordering of Korean stop phonemes and that tonal development is significantly related to the acquisition of Korean phonemic contrasts.

Keywords: stop contrasts; speech perception; F0

Introduction

Infants initially acquire human speech signals holistically as one meaningful chunk, then later come to realize that the signals 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 and decline non-native contrasts; hence, they learn to interpret speech sounds (Werker & Tees, 1984). As every spoken language has stop sounds in its consonant inventory, it is understandable that young children acquire stop sounds earlier compared to other consonant sounds in general (Jakobson, 1968). In the acquisition of stop sounds, phonetic features that are directly involved in characterizing stops are accepted while redundant features are declined in one's perceptual space. This process occurs early in language development and makes it possible to discriminate between native stop contrasts. Thus, the acquisition of stop

contrasts is directly linked to the development of relevant phonetic parameters that serve to distinguish stops.

The Korean stop system is considered typologically rare because it has three different voiceless stop contrasts that can be differentiated phonologically and phonetically: lenis, fortis, and aspirated. Due to the unusual typology of Korean stops, researchers have attempted to distinguish the three-way contrast using feature-geometrical differences in the underlying forms of the three stops (e.g., Han, 1992; Lombardi, 1991). Lenis stops are intervocalically voiced ([+ slack vocal folds]), and the actual phonetic differences between fortis and aspirated stops can presumably be predicted by the parameter of glottal width: [constricted glottis] for fortis stops, [spread glottis] for aspirated stops, and unspecified for lenis (Kochetov & Kang, 2017). These phonological representations correspond to the acoustic implementation of fortis and aspirated stops (Kagaya, 1974). These findings indicate that the feature specification of the three stops reflects the articulatory characteristics of Korean stop contrasts.

A number of phonetic studies have extensively analyzed the acoustic properties of the Korean stop system. Because these three stop contrasts are always realized as voiceless when positioned phrase-initially, both voice onset time (VOT) and fundamental frequency (F0) are used to distinguish among stop contrasts. It has been well established that VOT plays a primary role in discriminating among these stop categories, even though voicing is not involved in stop contrasts in Korean. Lenis and aspirated stops have long-lag VOT, while fortis stops have relatively shorter VOT, and the three categories overlaps in terms of VOT (Cho, 1996; Han & Weitzman, 1970; Hardcastle, 1973; Hirose, Lee, & Ushijima, 1974; C.-W. Kim, 1970; M.-R. Kim, 1994). As a secondary phonetic parameter, F0 contributes to identifying mainly lenis stops and does not necessarily serve to distinguish between fortis and aspirated stops (Cho, 1996; Han & Weitzman, 1970; Hardcastle, 1973; Kagaya, 1974; Kim, 1994). Phrase-initial lenis stops tend to lower the F0 of the following vowel, while fortis and aspirated stops tend to raise the F0 of the following vowel (Haudricout, 1954; Hombert, Ohala, & Ewan, 1979; House & Fairbanks, 1953).

Recently, however, Silva (2006), Wright (2007), and Kang (2014) proposed that F0, in addition to VOT, serves as a crucial cue for Korean stop distinction. According to these diachronic studies, Korean stop contrasts are undergoing tonogenesis, which is the process of sound change resulting in a loss of VOT differences and an increase in F0 differences between lenis and aspirated stops in young adults' speech. More specifically, fortis stops can still be distinguished by a distinctively short VOT, whereas articulatory distinction between lenis and aspirated stops requires a robust F0 difference. Among the three stop groups, lenis stops have the lowest F0, while aspirated stops have the highest F0.

The phonetic relationship between Korean stop contrasts in the dimension of VOT and F0 is well illustrated in Figure 1. The productions of stops by young adult speakers (mean age = 25.8 years) show that the considerable phonetic distance in VOT between fortis and non-fortis stop categories can generally be found across different speech styles. To distinguish between lenis and aspirated stops, F0 offsets the possible VOT overlap between them, while fortis stops have intermediate F0 values in any speech style.

Regarding the acquisition of stop sounds, many laboratory studies have concentrated on the discriminant ability in the perception of stops. One of the studies reported that 1- and 4-month-olds could discriminate between [ba] and [pa] in English using the phonetic parameter of VOT (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Several
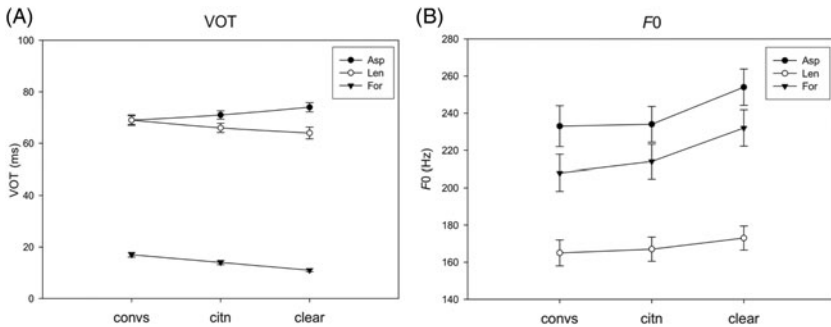
Figure 1. Mean values with standard errors for the production of Korean stops (aspirated, lenis, and fortis) in conversation, citation-form, and clear speech styles by young adult speakers for (A) VOT and (B) F0. Adapted from 'Clear speech production of Korean stops' by K.-H. Kang and S. G. Guion (2008), *Journal of the Acoustical Society of America*, *124*, p. 3913. Copyright 2008 by the Acoustical Society of America.

studies by Swingley and his colleagues also demonstrated that young children can discriminate the English stop contrast, although the primary goal of their research did not involve identifying infants' sensitivity to the stop contrast itself. For example, two studies by Swingley and Aslin (2000, 2002) tested children aged 14 to 23 months, exposing them to stimuli with a word-initial English stop sound, and their word recognition was influenced by the 'correct' onset sound. If the stop sound was replaced with another stop (e.g., changing *dog* to *tog*), it interfered with young children's word recognition. Using an eye-tracking method, Swingley, Pinto, and Fernald (1999) demonstrated that two-year-olds are able to distinguish English stops (the pair of *ball* and *doll*). Their findings indicated that, at the age of 24 months, children's perceptual representations are sufficiently established to allow them to discriminate their native phonetic contrasts and that the phonetic information is also stored with the lexical meanings of words in the word learning process. Notably, not all phonetic gestures are acquired simultaneously in early language development. Rather, relatively salient and familiar phonetic features are first involved in forming phonetic categories such that perceptual representations continuously undergo phonological attunement throughout development (Logan, 1992; Metsala, 1997; Swingley, 2009).

In the case of Korean, only a few phonetic-acoustic analyses of children's stop production in terms of VOT and F0 have been conducted (e.g., Kong, Beckman, & Edwards, 2011; Lee & Iverson, 2008). Some phonetic investigations have argued that the acquisition of fortis stops (which correspond to short-lag VOT stops) preceding the acquisition of lenis and aspirated stops should be understood as a language-universal acquisition pattern, since (pre)voiced or long-lag VOT stops are associated with articulation difficulties (Kim & Stole-Gammon, 2009; Kong *et al.*, 2011). 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 in air pressure between the supraglottal and the subglottal cavities should be made so that the vocal folds start to vibrate during the closure (Keating, Linker, & Huffman, 1983). The earliest emergence of fortis stops in children's production has been discussed in relation to VOT development for articulatory reasons. Regarding the non-fortis stops, the critical role of F0 has been shown in the distinction of lenis

from aspirated stops, but has not been dealt with in the context of tonogenesis (e.g., Kong *et al.*, 2011). Therefore, the two-dimensional dynamics have been suggested to account for the development of Korean stop contrasts.

With this prediction in mind, it is worth investigating how VOT or F0 development influences Korean toddlers' perceptual categorization of stops in a trading relation between the two phonetic parameters. It is assumed that production is guided by perceptual representations, since perceptual development precedes production in infants. However, previously reported linguistic research on the acquisition of Korean stop contrasts is biased towards the production side, and no perceptual analysis with a large body of data has been reported. Without a systematic assessment of toddlers' perception of stop contrasts, the ways in which multiple native phonetic parameters develop throughout phonological development to distinguish phonemic contrasts are unclear. Therefore, the current study focused on how the perceptual distinction of phonemic contrasts develops in toddlers and how such perceptual development represents toddlers' linguistic competence.

In this study, we designed a perception experiment assuming that, along with VOT, F0 necessarily develops during phonemic processing of Korean stop contrasts, which reflects the recent tonogenetic sound change that is currently under way in young adult speech; this means that F0 differentiation between lenis and aspirated stops will be amplified. We used two different sound stimuli in the experiment, natural stop sounds and synthesized ones that have an F0 continuum. By using this method, this study was able to observe Korean toddlers' perception pattern of Korean stops and to examine the effectiveness of the role of F0 in the perceptual distinction between lenis and aspirated stops. The study aimed to provide evidence to show how the developmental trajectories of VOT and F0 can affect the phonemic categorization and the mastery of ordering of contrastive phonemes in the Korean stop system.

Materials and methods

Participants

Fifty-eight children aged between 2;0 and 3;11 (years;months) were recruited for the perception experiment from two daycare centers in Seoul, Korea. None of the children had hearing or speaking disorders, and the children's parents or guardians were all Korean-language monolinguals. Among the 58 participants, the data of 10 children were not included in the analysis, either because they did not complete the task (8) or because they seemed unfocused (2), and thus their responses were judged to be unreliable despite their completion of the task. The remaining 48 child participants were divided into four different age-groups, as shown in Table 1.

Listening materials 1

The identification test was performed using a point-to-a-picture task. The 18 trials consisted of nine minimal pairs grouped by three different places of articulation (POA), which included every possible pair of lenis–fortis–aspirated homorganic stops. We decided to use only pairs in the task rather than triplets, which would have made the test more difficult, as choosing one out of three alternatives might be

Table 1. Child participant information

Age-group	M	SD	Male	Female	Total
2;0–2;6	2;3.21	0;2	5	6	11
2;7–3;0	2;9.18	0;1.18	4	4	8
3;1–3;6	3;3.3	0.1;12	8	7	15
3;7–3;11	3;8.27	0;1.3	3	11	14

Table 2. The pairs and words used in the perception test. The words highlighted in bold are not listed in MCDI-K.

POA and stop category	Pairs	Word	English gloss	Word	English gloss
Labial lenis–fortis	/p/–/p'/	/paŋ/	room	/p'aŋ/	bread
Labial lenis–aspirated	/p/–/pʰ/	/pal/	foot	/pʰal/	arm
Labial fortis–aspirated	/p'/–/pʰ/	**/p'ʊl/**	horn	/pʰʊl/	grass
Alveolar lenis–fortis	/t/–/t'/	/tal/	moon	**/t'al/**	daughter
Alveolar lenis–aspirated	/t/–/tʰ/	**/tɔk'i/**	hammer	/tʰɔk'i/	rabbit
Alveolar fortis–aspirated	/t'/–/tʰ/	/t'ʌk/	rice cake	/tʰʌk/	chin
Velar lenis–fortis	/k/–/k'/	**/kʊl/**	cave	**/k'ʊl/**	honey
Velar lenis–aspirated	/k/–/kʰ/	/kɔŋ/	ball	/kʰɔŋ/	beans
Velar fortis–aspirated	/k'/–/kʰ/	/k'inʊn/	wearing	**/kʰinʊn/**	turning on

too difficult for young children at the target age. Table 2 presents the wordlist used for the experiment. It was impossible to find minimal pairs to assess the target-aged children from the list of the MacArthur Communicative Development Inventory in Korean (MCDI-K; Pae, Chang, Kwak, Sung, & Sim, 2004); therefore, we used some words (shown in grey cells in Table 2) that were not included in MCDI-K. The sound stimuli used were the natural production of a female Korean-language monolingual speaker (age = 30) who lived in Seoul, Korea.

Each pair was presented twice in random order. Before the beginning of the experiment, each participant played with the experimenter using the pictures (and other toys) to ensure that the participant was aware of what each picture depicted. To confirm that the child participants were aware of what each picture depicts, they were encouraged during playtime to pronounce the words while viewing each picture. The two pairs, including /pal/–/pʰal/ and /kɔŋ/–/kʰɔŋ/ pairs, were played only once (for /pal/ and /kɔŋ/, respectively) because these pairs were also used for synthesized stimuli tests. Thus, 16 pairs were used for natural stimuli.

Listening materials 2

The production of /pal/ 'foot' and /kɔŋ/ 'ball' by the same female speaker was used for the synthesis. Using the pitch synthesis function in Praat (Boersma & Weenink, 2015), the onset of the following vowel was manipulated to have a 15 Hz difference

Table 3. Synthesized stimuli used in the perception experiment

	Natural production	Synthesized stimuli				
F0 at vowel onset of /pal/	≈ 200 Hz	215 Hz	230 Hz	245 Hz	260 Hz	275 Hz
F0 at vowel onset of /kɔŋ/	≈ 220 Hz	235 Hz	250 Hz	265 Hz	280 Hz	295 Hz
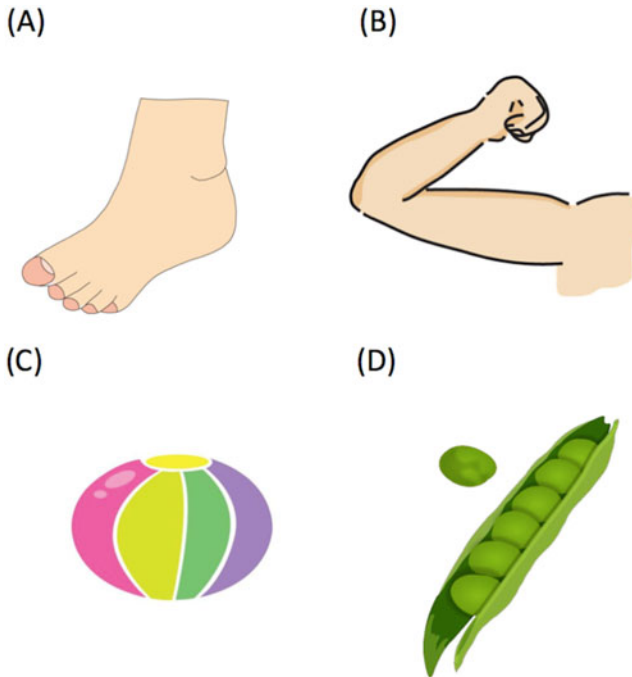
Figure 2. Two paired sets in the perception test. /pal/ (A)–/pʰal/ (B), and /kɔŋ/ (C)–/kʰɔŋ/ (D).

between the tokens. As shown in Table 3, the two monosyllabic words were manipulated to have six different F0s at vowel onset with fixed VOT. The sound stimulus with the lowest F0 represented the average natural production of each word by the speaker. The vowel onset of /pal/ was manipulated to 215 Hz, 230 Hz, 245 Hz, 260 Hz, and 275 Hz with fixed VOT (70 ms). The vowel onset of /kɔŋ/ was manipulated to 235 Hz, 250 Hz, 265 Hz, 280 Hz, and 295 Hz with fixed VOT (75 ms). As the F0 values at vowel onset changed, the consecutive pitch points also changed so that the original pitch contour was nearly maintained, resulting in stimuli that sounded close to natural. The paired pictures shown in Figure 2 were used.

Procedure

The pictures were presented one pair at a time to each child participant on a personal laptop. The participant was asked to point to one of the two pictures in a given pair immediately 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 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 natural productions and synthesized sounds. The perception experiment was conducted in a quiet room in the library of a daycare center. Each session took approximately 10 minutes and consisted of 31 trials (5 fillers + 10 synthesized stimuli + 16 natural stimuli). The two different kinds of stimuli were presented in random order and used in the same session for the identification task. The main reason for combining the two types of stimuli was that it was assumed that toddlers would be unable to maintain their concentration for two separate sessions. Moreover, the two different types of stimuli needed fillers; thus, combining them was the considered the best option, as they functioned as natural fillers for one another. The five words that were used only as fillers were bi- or trisyllabic words. If a child seemed distracted during a question, the same question was played again by the experimenter. Every answer given by the participant was documented by the experimenter during the test.

Results

Natural stimuli

During the session, 768 responses (16 trials × 48 child participants) were collected. Overall, the results show that the child participants identified the given natural sounds with 86.7% accuracy. The accuracy level varied depending on which stop category was given: aspirated stops were most correctly perceived (90.1%), followed by fortis stops (89.9%), while lenis stops were the least accurately identified (80.9%) by the children.

The accuracy levels for each trial are presented in Table 4. The accuracy levels for the identification of the same target phoneme differed depending on what minimal pairs were provided for the choice. For example, the 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 /tʰ/ as their choices, only 73% of the responses identified /t/ correctly.

Children generally identified natural sounds well when fortis and one of the non-fortis stops (i.e., lenis or aspirated) were paired as the given alternatives. However, when the two non-fortis stops were paired, accuracy levels decreased. For instance, when the toddlers were provided with velar lenis /k/ as a target sound and had to decide between /k/ and velar aspirated /kʰ/, their accuracy rate was the lowest (66.7%). Velar fortis /k'/ was the easiest phoneme to correctly identify when it was provided with lenis stop /k/ (100%). This tendency was also found in the identification of /p/. Children identified /p/ well when it was paired with /p'/ (89.6%), while the same stop phoneme /p/ was 77.1% correctly identified when paired with /pʰ/.

Figure 3 shows differences in accuracy between age-groups. Overall, children's perceptual accuracy increased with age. The general perceptual patterns were generally maintained throughout the target ages. The estimated perceptual accuracy for fortis stops and aspirated stops was higher than that for lenis stops for all age-groups. Across all age-groups, the perceptual accuracy for lenis stops was always the lowest.

Table 4. Degree of accuracy of each trial in the perception test with natural stimuli

| Answer	Used pairs as alternatives	# Correct answers/# Total responses	Degree of accuracy (%)
p	/paŋ/–/p'aŋ/	43/48	89.6
p'	/paŋ/–/p'aŋ/	44/48	91.7
p	/pal/–/pʰal/	37/48	77.1
p'	/p'ʊl/–/pʰʊl/	40/48	83.3
pʰ	/p'ʊl/–/pʰʊl/	45/48	93.8
t	/tal/–/t'al/	44/48	91.7
t'	/tal/–/t'al/	40/48	83.3
t	/tɔk'i/–/tʰɔk'i/	35/48	73.0
tʰ	/tɔk'i/–/tʰɔk'i/	45/48	93.8
t'	/t'ʌk/–/tʰʌk/	46/48	95.8
tʰ	/t'ʌk/–/tʰʌk/	46/48	95.8
k	/kʊl/–/k'ʊl/	42/48	87.5
k'	/kʊl/–/k'ʊl/	48/48	100
k	/kɔŋ/–/kʰɔŋ/	32/48	66.7
k'	/k'inʊn/–/kʰinʊn/	41/48	85.4
kʰ	/k'inʊn/–/kʰinʊn/	37/48	77.1
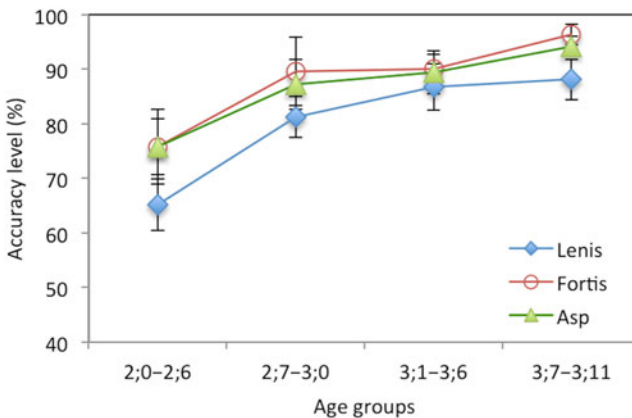
Figure 3. Accuracy-level differences by age-group across three stop categories. Error bars represent standard errors.

To confirm the effect of predictor variables, including stop categories and children's age, on the observed perceptual accuracy of the toddlers, a mixed-effects logistic regression model (Raundenbush & Bryk, 2002; Snijder & Bosker, 1999) was designed and conducted. The *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) was implemented in R (R Development Core Team, 2011), and the *glmer* function was

Table 5. Output of the mixed-effects logistic regression model for the effect of age on children's correct perception of stop contrasts, with lenis as the reference category.

Random effects:			
Groups	Name	Variance	Std. dev.
child_id	(intercept)	2.652e-17	5.15e-09
Number of obs.: 768, groups: child_id, 48			
Fixed effects:			
	Estimate	Std. error	z value	Pr (>|z|)
(Intercept)	−2.00	0.69	−3.30	<0.001***
Age	0.10	0.02	5.66	<0.001***
Fortis	0.63	0.24	2.59	0.01**
Asp	1.07	0.31	3.41	0.001***

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.
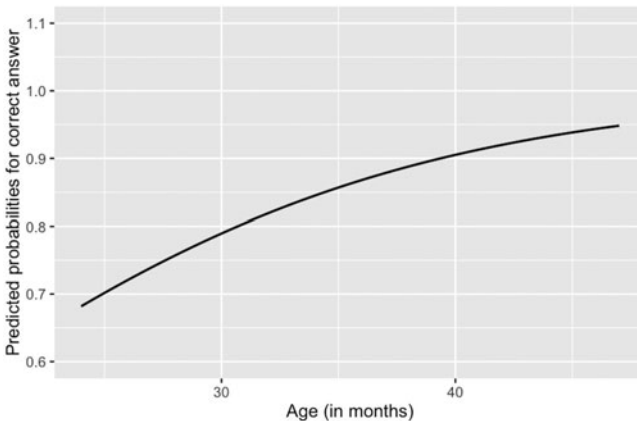
Figure 4. Relationship between children's age and correct answers depicted with a logistic curve for all responses.

used. A model that demonstrates the effect of age on the accurate perception of each individual with the three phonation types (lenis, fortis, and aspirated) was built and tested. This model predicted correct/incorrect answers as dependent variable, and the log odds of the probability of being correct were predicted by two parameters (stop category and age). Lenis is the reference category, and by-listener adjustment to the intercept was a random effect.

As shown in Table 5, the effects of age, lenis, fortis, and aspirated stops were significant using this statistical model (p-values < .05). The fixed-effects predictors, age, fortis, and aspirated, had positive coefficients, which means that these predictors are positively related to the probabilities of correct answers in the perception experiment; that is, children's perceptual accuracy increases with age, and the perception of fortis and

Table 6. Output of the mixed-effects logistic regression model for the effect of the interaction between children's age (in months) and stop category on children's correct responses, with lenis as the reference category.

Random effects:			
Groups	Name	Variance	Std. dev.
child id	(intercept)	3.294e-13	5.74e-07
Number of obs.: 768, groups: child id, 48			
Fixed effects:			
	Estimate	Std. Error	z value	Pr (> |z|)
---	---	---	---	---
(Intercept)	−1.47	0.83	−1.77	0.08
Age	0.08	0.02	3.47	<0.001***
Fortis	−0.65	1.32	−0.49	0.62
Asp	0.46	1.67	0.28	0.78
Age:Fortis	0.04	0.04	1.00	0.32
Age:Asp	0.02	0.05	0.37	0.71

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

aspirated stops is more accurate than that of lenis stops. As Figure 4 illustrates, children's age (in months) was significantly positively related to the probabilities of correct responses; thus, from the age of two to four years, a significant perceptual development in stop acquisition can be observed.

To predict the interaction of stop category and age on the children's responses, another mixed-effects regression was conducted. This time, the interaction term (stop category × age) was added to the former model to predict the effect of children's age on the correct perception of each stop category. Lenis was the reference category. As represented in Table 6, the fixed-effects coefficients showed that most of the predictors, except for age ($p < .001$), did not have significant effects on correct identification by the children (p-values > .05). Here, no significant correlation between children's age and the perceptual accuracy of fortis or aspirated stops could be found ($p > .05$), but the interaction between age and correct perception of lenis stops was significant. That is, with an increase in age (in months), there is an 8% increase (*coef.* = .08) in the odds of correct answers for lenis stops. According to the results, the perception of lenis stops differed significantly from two to four years of age. As presented in Figure 3, perceptual accuracy of fortis and aspirated stops did not change drastically over age.

As the toddlers were consistently the least successful in identifying lenis stops compared to the other stop types, we attempted to analyze the context in which they would have experienced more difficulties. Table 7 and Figure 5 illustrate changes in perceptual accuracy for lenis stops across age-groups. With fortis stops, the second age-group identified almost all lenis stops perfectly (95.8%). However, with aspirated stops, the oldest age-group showed an 81% accuracy level in the perception of lenis stops, with accuracy levels consistently increasing with age.

Table 7. Correct responses in the perception of lenis stops according to age-groups. The numbers in parentheses represent the total number of responses acquired.

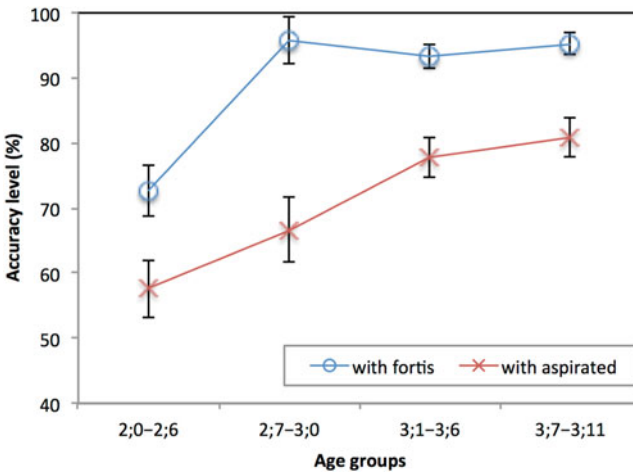|	Age-groups			
	2;0–2;6 (N = 33)	2;7–3;0 (N = 24)	3;1–3;6 (N = 45)	3;7–3;11 (N = 42)
With fortis	24	23	42	40
With aspirated	19	16	35	34

Figure 5. Comparison of accuracy levels in the perceptual identification of lenis stops by age-group when fortis and aspirated minimal pairs are provided. Error bars represent standard errors.

Hence, the perception of lenis stops showed consistently lower scores with aspirated stops than with fortis stops. Since not all items used for the experiment were listed in MCDI-K, we attempted to determine whether the predictor of the relative difference of lexical familiarity between the given alternatives might significantly affect this biased result. The words listed in MCDI-K were treated as high-familiarity words, whereas those not listed in MCDI-K were considered low-familiarity words. The possible effect of word familiarity difference on the perception of lenis stops was examined using another mixed-effects logistic regression. This model was tested with factors of a paired stop category (aspirated or fortis stop) and word familiarity difference between the paired alternatives (higher/lower lexical familiarity of lenis stops, or no lexical familiarity difference). This model predicted the correct vs. incorrect answers as the dependent variable, and the individual participant was the random effect varying an intercept. In Table 2, the words highlighted in grey that were paired with lenis were treated as low-familiarity words (/t'al/, /tɔki/, /kʊl/, and /k'ʊl/), while the others were all treated as high-familiarity words. In the case of lenis–fortis pairs, one pair (/tal/ 'moon' –/t'al/ 'daughter') differed in lexical familiarity; the lenis has higher lexical familiarity. In the case of lenis–aspirated pairs, one pair (/tɔk'i/
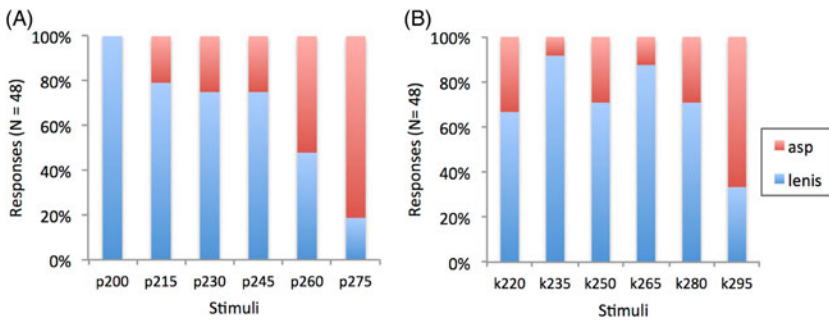
Figure 6. Results of the identification of synthesized /p/ (A) and /k/ (B) by toddlers.

'hammer' – /tʰɔk'i/ 'rabbit') differed in lexical familiarity, and the lenis had lower familiarity. The statistical results revealed a significant effect of a paired stop category, showing that when fortis stops were paired, the perception of lenis stops tended to be significantly correct ($p < .01$). However, no other significant effect of lexical familiarity difference on the perception of lenis stops was found (p-values > .1).

In addition, no effect of lexical familiarity was found on the results in general. A logistic regression with a factor of word familiarity of the target word (high or low) showed no significant effect of the factor on the correct perception of three stop types ($p > .1$). Regarding an interaction term between children's age (in months) and word familiarity of the target word (high and low), an ANOVA confirmed that a fixed effect of the interaction between children's age and word familiarity on the correct perception was not significant ($X^2 = 3.77$, $df = 1$, $p > .05$).

Synthesized stimuli

As shown in Figure 6, the synthesized set of /pal/–/pʰal/ shows that the children's lenis–aspirated stop distinction differed according to a substantial F0 difference between lenis /p/ and aspirated /pʰ/. The lenis /p/ with 200 Hz at its vowel onset, which is considered an average lenis articulation, was identified as /p/. In contrast, the synthesized /pal/ with 275 Hz at its vowel onset was perceived as aspirated /pʰ/ in 81.3% of the responses. The synthesized /p/ stimuli with 215 Hz, 230 Hz, and 245 Hz tended to be perceived as lenis /p/, but an F0 difference of 60 Hz or greater meant that the stimuli were more likely to be identified as the aspirated counterpart /pʰ/.

The results of the velar stop contrast /k/–/kʰ/ did not correspond to those of their labial counterparts. Overall, the responses did not change based on the relative difference in F0 for the /k/–/kʰ/ distinction. The children's responses fluctuated; 92% of the responses were /k/ in the case of the 235 Hz stimulus while 70% of the responses were /k/ in the case of 280 Hz stimulus. The only case in which more than 50% of the cases were identified as aspirated /kʰ/ is the stimulus presented at 295 Hz.

The effect of age or F0 differences on the children's responses was analyzed using the mixed-effects logistic regression model with all of the obtained responses to both /p/ and /k/ stimuli. Using this model, the log odds of the probability of being perceived as aspirated (i.e., /pʰ/ and /kʰ/) were predicted by two independent factors (age and F0 difference). Each child participant was the random effect varying an intercept. The regression results are presented in Table 8 and Figure 7.

Table 8. Output of the mixed-effects logistic regression model for the effect of predictors on the probabilities for aspirated stops

Random effects:			
Groups	Name	Variance	Std. dev.
child id	(Intercept)	0.368	0.6067

Number of obs.: 576, groups: child id, 48

Fixed effects:			
	Estimate	Std. error	z value	Pr (>|t|)
(Intercept)	−1.61	0.78	−2.06	0.04 *
Age	−0.02	0.02	−1.02	0.31
F0_difference	0.04	0.01	8.52	<0.001***

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.
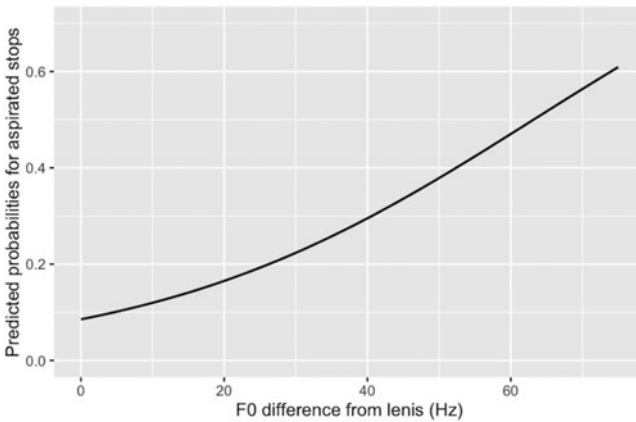
Figure 7. Relationship between F0 differences from lenis stops and the perception of aspirated stops depicted with a logistic curve.

F0 differences had a significant effect on the probability of being perceived as an aspirated stop for stimuli ($p < .001$). The F0 difference for each stimulus significantly affected the toddlers' responses to both series of /p/ and /k/. As the F0 difference increases, the odds of the stimuli being identified as aspirated stops increases by 3.8% (*coef.* = .038). Age did not influence the identification of aspirated stops ($p > .1$).

Discussion

The perception experiment with natural sound stimuli showed that perceptual accuracy is lowest in the identification of lenis stops, and that, in general, this perceptual pattern is applicable to all four age-groups. Young children aged two to four years seem to be able to correctly perceive fortis and aspirated stops better than lenis stops. In particular,

the perception of lenis stops was the least successful when they were paired with aspirated stops as alternatives. This consistent observation directly indicates that lenis stops are perceptually less salient compared to fortis or aspirated stops. Overall, it is assumed that the perceptual system surrounding those phonation types has not yet stabilized to the level of an adult speaker, since the perceptual accuracy of the child participants was not close to 100%.

Although this study's findings point to the imperfect perceptual abilities of children in the multi-parametric acoustic dimension, the interpretation of the results should also consider that the toddler participants had to identify the stimuli in an experimental setting in which contextual information about the target phonemes was extremely restricted. In the actual session, they may have depended on lexical familiarity or other lexical properties when the perceived acoustic difference was not apparent to them, although they were given enough time to become familiar with each picture with the words pronounced. For example, the pair of velar fortis (/k'/) and velar aspirated (/kʰ/) stops consisted of verb-inflected forms, while other pairs consisted of only noun forms. The average perceptual accuracy for velar fortis and aspirated stops (81.3%) was somewhat lower than that for bilabial or alveolar fortis and aspirated pairs (88.6% and 95.8%, respectively). We cannot rule out the possibility that the children's lexicon and the word recognition processes for nouns and verb-inflected forms could have affected their responses, although velar consonants are usually acquired later compared to labial or alveolar ones (Jacobson, 1968). Another concern involved differences between the given alternatives in lexical familiarity, since children tend to encode familiar words with phonetic details more easily (e.g., Swingley & Aslin, 2000). In particular, lenis stops were less successfully perceived when they were paired with aspirated rather than fortis stops. The lowest perceptual accuracy of lenis stops might have been caused by an imbalance in lexical familiarity between the minimal pairs. We used statistical modeling to determine the effect of word familiarity on the correct perception of lenis stops in this experiment, which resulted in no significant effect of lexical familiarity. In general, no other significant correlation was found between lexical familiarity and other predictors, including children's age and stop categories. This might be because the word familiarity of the target words was not carefully controlled in the first place, so the wordlist had a relatively small number of low-frequency words (six out of 18 target words). For more accurate analysis of the effect of word familiarity, further investigation is required.

Although these kinds of interfering factors in the analysis may make the results seem inconclusive, this study consistently showed that when lenis stops were the target with aspirated counterparts, the perceptual accuracy was similarly low across the three places of articulation (77.1% for /p/, 73% for /t/, and 66.7% for /k/). Of course, this forced-choice selection method might not be the best way to capture the accuracy of children's phonological representations in perceiving contrastive phonemes. However, this task still makes it possible to draw a clear outline of the relationship between the three different laryngeal contrasts and their relative perceptual salience in the multidimensional acoustic space. Toddlers' responses suggest that their perceptual mapping functions are distinctively based on high-pitched phonemes. It was observed that fortis and aspirated stops tend to be perceived more accurately than lenis stops, indicating that high-pitched stop categories in two extreme VOT regions are well specified in children's perceptual space. Children had difficulty perceiving lenis stops, despite their distinctively low F0.

In addition, assuming that the perceptual space of children of the target age is not sufficiently developed to differentiate F0 differences and to categorize the lower range of F0, the worst performance in identifying lenis stops between the pair of lenis and aspirated stops can be well explained. As shown in the logistic mixed-effects analysis in Table 6, only the perception of lenis stops showed a significant change in perceptual accuracy with age. This finding indicates that, at the age of two to four years, a significant linguistic development in the perception of lenis stops occurs, which implies phonemic development in the F0 dimension. Some other findings have also verified this hypothesis, showing that the perception of lenis stops became poorer when they were minimal-paired with aspirated counterparts rather than with fortis counterparts. This perception pattern suggests that the same phoneme can be perceived differently depending on what phonetic parameter is predominantly involved in the perception.

The results of the perception test with natural sound stimuli imply that around two to four years of age, children's perception system for Korean stop contrasts develops in the VOT dimension, and they are able to distinguish VOT differences in high-pitched stop types, since the perception of fortis and aspirated stops is relatively more successful. The lowest perceptual accuracy with lenis stops indicates that lenis stops need another acoustic dimensional categorization to be identified in their perceptual space, and that, at the age of two to four years, children may not be sensitive enough to low F0 values to establish perceptual categorization of lenis stops.

As tonogenetic sound change has been reported in Seoul Korean (Kang, 2014, among others), F0 has become the most informative acoustic cue to distinguish between lenis and aspirated stops because of the phonetic merger of VOT between the two categories. To investigate the development of F0 as a contrasting cue in perceptual acoustic space, the same children were presented with sound stimuli synthesized to have various F0 onsets. The significant effect of F0 difference between lenis and aspirated stops has been found in the perception of aspirated stops.

Unsurprisingly, individual variations in children's responses were observed, since not all children provided stable responses. For example, one child identified a 215 Hz stimulus as aspirated /pʰ/ but identified the 245 Hz stimulus as lenis /p/. Each stimulus was tested once, so it was impossible to calculate the average response per stimulus. To make judgments about the stability of responses, 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 an AIC greater than 10.0 would not be acceptable as a stable response; thus, only the responses with an AIC under 10.0 were considered stable.[1] Therefore, in the current study, if a child's confusion persisted over two steps of trials, we did not consider it a valid response. If the child's choice was reversed within two steps, it was considered a stable response.[2] Table 9 shows the number of children in each age-group who showed response stability. In general, the stable response rate gradually increases according to age. Although the velar stop pair showed poorer and less consistent perceptual patterns, the response stability increased with the participants' ages,

---

[1]The AIC was 4.0, but a perfect reversal – as in 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*).

[2]If a child responded 0 0 0 1 0 1 to 200 Hz, 215 Hz, 230 Hz, 245 Hz, 260 Hz, and 275 Hz, respectively, this was considered a stable response. If a child chose 0 0 1 0 0 1, this was not considered a valid response.

Table 9. Number of subjects and stable responders 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 in /p/	7 (63.6%)	7 (87.5%)	13 (86.7%)	14 (100%)
Number of stable responders in /k/	5 (45.5%)	3 (37.5%)	11 (73.3%)	11 (78.6%)

suggesting that children's perceptual accuracy increases with age according to the development of F0.

Children's perceptual thresholds for the phonemic categorization of aspirated stops varied relative to their different F0 development stages, which are accordingly affected by their age. The response stability increases as children grow up, while children's confusion in identifying lenis and aspirated stops is retained to some extent until three years of age. Despite some differences in response stability, it seems that phonetic attunement to F0 occurs at the age of two to four years, and perfect categorical perception in the F0 dimension is not found until four years of age. One of the most important findings here is that toddlers under four years of age still need to undergo further phonemic development with fine phonetic details. Therefore, at a certain stage of development, perceptual phonemic categorization should stabilize with the phonetic norms for each category. Establishing their phonetic standards, the children began to distinguish a lenis–aspirated contrast. Through this reorganization of phonemic stop categories in the development of the F0 dimension, children were able to correctly and accurately perceive two different phonation types.

Recall, however, that the experiment with the two sets of tested stimuli (bilabial and velar) showed differences in output. That is, the perception of synthetic velar stops fluctuated more compared to that of bilabial stimuli. The irregular perception patterns in the case of /k/ stimuli can be explained in light of the intrinsic phonetic differences between the two series of stops arising for physiological reasons; velar stops have closure at a more posterior part of the vocal tract, whereas bilabial stops have closure at the lips. This articulatory difference causes inherent phonetic variations in VOT, and it has been well understood that velar stops usually have longer VOT compared to labial stops (Cho & Ladefoged, 1999, among others). Accordingly, the unparalleled results in the two different places of articulation could have emerged because language-learning child listeners might be less sensitive to relative F0 differences in the case of velar stops, since those stops have relatively stronger and longer VOTs. Along the same lines, bilabial stops usually have the shortest VOT among the Korean stop sounds, and the phonetic compensation in perceiving them might occur in the dimension of F0, as listeners might be more sensitive to F0 in perceiving bilabial stops. In the perceptual development of labial stop contrasts, it is likely that labial stops require less difference in F0 compared to the other stops. The role of VOT in identifying the POA of stop sounds has not been highlighted in the previous literature, since the formant transitions are more informative cues to differentiate the perceived place of articulation (Stevens & Klatt, 1974, among others). However, the development of Korean stop contrasts requires two parametric controls, and since the dynamic interaction of those two parameters throughout development is still unclear, it might be possible that a certain property

of one dimension can affect the relative sensitivity to the other dimension (e.g., Shuai & Gong, 2012; Winn, Chatterjee, & Idsardi, 2013). In addition, the acquisition of labial stops well ahead of that of velar stops has been considered a general pattern in children's language development (Jakobson, 1968). These speculations help explain some of the discrepancies in children's perception as a developmental process of phonemic distinction. The effect of lexical items used in the experiment was hardly noticeable, since the words in the velar set ('ball' and 'beans') seem familiar enough for the children to distinguish. The two words in the set are listed in the MCDI-K.

This study has shown that the acquisition order among native phonological contrasts is directly related to the perceptually salient phonetic parameter. The acquisition of fortis and aspirated stops occurs before the acquisition of lenis stops, suggesting that Korean toddlers' perceptual representations develop to categorize VOT variations in a high-pitched region at an earlier stage of learning. This finding also suggests that toddlers do not yet recognize categorical phonetic differences in the F0 dimension as accurately as in the VOT dimension.

These findings suggest that VOT functions as an effective tool in relation to another phonetic parameter, F0, in Korean toddlers' perceptual space, resulting in a certain mastery of ordering among stop contrasts. Historically, VOT has been a successful measure of the aerodynamic mechanism that pertains to phonetic stop differentiation, and it is a language-universal tool for distinguishing stop categories. The early acquisition of fortis stops has been understood such that short-lag stops are acquired earlier than long-lag or lead voiced stops (Kewley-Port & Preston, 1974). This is because articulating short-lag VOT stops is less demanding compared to articulating the other two types of VOT stops, which require precise temporal control between the glottis opening and the oral constriction release. Of course, this early acquisition of short-lag VOT is not a sufficient explanation for the mastery of ordering of Korean stop contrasts, since another parametric control, F0, is involved, which has been amplified following a tonogenetic sound change in Seoul Korean.

In the acquisition of tone languages, lexical tonal features are acquired earlier than segmental features. Acquisition research on tone languages, such as Thai and Mandarin Chinese, uniformly reports that tone acquisition occurs prior to the full development of segmental features (Li & Thompson, 1977; Tse, 1978; Tuaycharoen, 1977). Infants' sensitive perception of tonal features in speech (Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989) and in music (Krumhansl & Jusczyk, 1990; Trainor & Trehub, 1992) has been consistently observed, and even in a non-tone language, such as Japanese, infants up to approximately 18 months of age tend to treat lexical pitch variation as an informative cue to word recognition (Ota, Yamane, & Mazuka, 2018). Thus, the findings on the development of lexical items with suprasegmental phonetic features, such as pitch and duration, indicate that infants can incorporate F0 variation into lexical information and use it as a contrasting cue to differentiate lexical items until they establish perceptual reorganization with the development of segmental features (Hay, Graf Estes, Wang, & Saffran, 2015; Singh & Foong, 2012; Singh, Hui, Chan, & Golinkoff, 2014; Singh, Poh, & Fu, 2016). Infants' perceptual sensitivity to F0 decreases depending on the sound input, so it is generally assumed that toddlers acquire F0 if it acts as a lexically contrasting cue before a segmental feature: VOT. Therefore, the reason that the acquisition of F0 does not occur prior to that of VOT in the development of Korean stop contrasts is puzzling.

The developmental pattern of VOT and F0 in the Korean stop system is considered language-specific because the phonetic feature with VOT differentiation is classified as

segmental phonology, while pitch difference can be classified as a lexical tonal feature. This language-specific acquisition pattern might be a very complex issue that cannot be resolved by conducting a simple experiment. A few perception studies with Korean adult speakers have provided empirical evidence that VOT is perceptually more salient than F0 in the distinction between Korean stop contrasts. Most recently, Kong and Lee (2018) suggested that increased F0 significantly affects the perceptual distinction of aspirated stops from lenis stops, but adult listeners generally showed greater perceptual dependency on VOT over F0 in perceiving Korean stop categories. However, research on children's perceptual bias for VOT over F0 has not been conducted in the Korean context.

One possible explanation for this acquisition pattern is that, in a phrase-initial position, a lowered F0 on the vowel onset would be the main phonetic cue for defining lenis stops, and F0 variations for phrase-medial lenis stops are rarely observed due to the overall phrasal intonation (Cho, Jun, & Ladefoged, 2002; Jun, 1996). Lexical pitch variations can be easily overlaid by the pitch movements at a higher prosodic unit in Korean. Lenis stops can be realized high-pitched if fortis or aspirated stops precede lenis stops in the same intonational phrase. Along the same lines, in a low-pitch environment, the F0 contour very slightly increases following an obstruent, and a conflict occurs between gestural and segmental tonal features (e.g., Hanson, 2009). These unmatched pitch patterns across different prosodic levels might affect the acquisition of F0 as a lexically contrasting cue in earlier stages of development. Even in the acquisition of a tone language, such as Mandarin Chinese, children have difficulty identifying certain tonal differences, since intonational pitch patterns do not correspond with lexical tonal patterns until four to five years of age (Singh & Chee, 2016). Thus, the acquisition of phonemes in the F0 dimension involves understanding multilevel prosodic structures and incorporating multilayer phonetic rules. Accordingly, establishing an F0 dimensional distinction might be delayed compared to the categorical perception in the VOT dimension.

Another possible account for this acquisition pattern is that Seoul Korean is undergoing tonogenesis and acquiring tonal contrasts even though Korean is not historically a tone language. As reported, a tonal contrast has recently begun to develop in the Korean stop system; therefore, the roles of F0 and VOT in the discrimination of stop contrasts are changing (Kang, 2014, among others). As this sound change has not yet been fully developed and established, children's language input is not constant. Most children's parents belong to a young generation whose production of stop contrasts enhanced F0 differentiation; however, whether these children's parents are their primary caregivers is unknown, and whether the use of phonetic cues in child-directed speech reflects the tonogenetic sound change is unclear. Recent research on child-directed speech revealed that Korean mothers enhance VOT differences between lenis and aspirated stops at an early stage of their children's language development while having a significant F0 difference in adult-directed speech. When the children are at the multiword stage, a significant F0 differentiation between lenis and aspirated stops is found in child-directed speech (Ko, 2018). To prove the effects of sound input on the perceptual development of children, further research should be conducted. The experiment in this study evaluated children's responses to only synthetic F0 values, not varying VOT values; hence, we were unable to compare the role of the two phonetic parameters in the distinction of aspirated stops from lenis stops. The logistic regression analysis indicated that F0 differences significantly affect the perception of aspirated stops

from lenis stops, but whether VOT changes also influence children's perceptual distinction between lenis and aspirated stops is still unclear. If a sound change is currently under way, it is difficult to examine the relationship between sound input and the developmental trajectories of phonetic parameters, but it is possible that inconstant sound input affects the delayed development of a certain phonetic cue. A certain degree of VOT differentiation between lenis and aspirated stops in the speech of conservative speakers might defer children's phonological specification in the F0 dimension.

Therefore, it is important to monitor the ways in which Korean toddlers' acquisition patterns continue to evolve to uncover the acquisition ordering of segmental and suprasegmental features. The perceptual salience of VOT or F0 in the development of Korean stop contrasts should be reconsidered in the context of tonogenesis, and more sensitive methods should be designed to assess children's perceptual accuracy.

Conclusion

The progress of linguistic development in early childhood depends on speech perception. Despite the importance of speech perception, no systematic perception research on Korean stop contrasts has been conducted. In addition, the development of tonal contrasts in Standard Korean has increased the role of F0 as a determinant in distinguishing between stop contrasts. It is worth investigating how this tonogenetic process might be related to young children's perceptual development of F0.

The findings from the experiment indicate that the perceptual identification of high-pitched stop categories, fortis and aspirated, depends on VOT differences, while F0 differences are affecting primarily the perceptual distinction between lenis and aspirated stops. This study attempted to show how Korean toddlers' perceptual development occurs in two-dimensional space, as the Korean language develops tonal contrasts in the native stop system. Since no perception study with young children aged from 24 months to 47 months has been reported, this study provides a piece of empirical evidence to trace the development of the two phonetic parameters in the acquisition of stop contrasts.

References

Bates, D., Maechler, M., Bolker, B., & Walkers, S. (2015). Fitting linear mixed-effects models using lme4. *Jounral of Statistical Software*, *67*(1), 1–48.

Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer* [Computer program]. Version 5.4.18. Online <http://www.fon.hum.uva.nl/praat/>.

Cho, T. (1996). *Vowel correlates to consonant phonation: an acoustic-perceptual study of Korean obstruents* (unpublished master's thesis), University of Texas at Arlington.

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., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*, 207–29.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–6.

Han, J.-I. (1992). On the Korean tensed consonants and tensification. *CLS*, 28, 206–23.

Han, M. S., & Weitzman, R. S. (1970). Acoustic features of Korean /P,T,K/, /p,t,k/ and /pʰ,tʰ,kʰ/. *Phonetica*, 22, 112–28.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal of the Acoustical Society of America*, 125, 425–41.

Hardcastle, W. J. (1973). Some observations on the tense–lax distinction in initial stops in Korean. *Journal of Phonetics*, 1, 263–72.

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

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: the contrastive use of lexical tone in early word learning. *Child Development*, 86, 10–22.

Hirose, H., Lee, C.-Y., & Ushijima, T. (1974). Laryngeal control in Korean stop production. *Journal of Phonetics*, 2, 145–52.

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, 105–35.

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

Jun, S.-A. (1996). Influence of microprosody on macroprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics*, 92, 97–116.

Kagaya, R. (1974). A fiberscopic and acoustic study of Korean stops, affricates, and fricatives. *Journal of Phonetics*, 2, 161–80.

Kang, K.-H., & Guion, S. G. (2008). Clear speech production of Korean stops. *Journal of the Acoustical Society of America*, 124, 3909–17.

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.

Keating, P. A., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, 11, 277–90.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16, 66–8.

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. (1970). A theory of aspiration. *Phonetica*, 21, 107–16.

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

Kim, M., & Stoel-Gommon, C. (2009). The acquisition of Korean word-initial stops. *Journal of Acoustical Society of America*, 125(6), 3950–61.

Ko, E.-S. (2018). Mothers would rather speak clearly than spread innovation: the case of Korean VOT. In *Proceedings of the 1st Hanyang international symposium on Phonetics and cognitive sciences of language* (pp. 30–1). Seoul: Hanyang Institute of for Phonetics and Cognitive Sciences of Language.

Kochetov, A., & Kang, Y. (2017). Supralaryngeal implementation of length and laryngeal contrasts in Japanese and Korean. *Canadian Journal of Linguistics / Revue canadienne de linguistique*, 62(1), 1–38.

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.

Kong, E. J., & Lee, H. (2018). Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech*, 61(3), 384–408.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, 1, 70–3.

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

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4, 185–99.

Logan, J. S. (1992). *A computational analysis of young children's lexicons* (Unpublished doctoral dissertation), Indiana University.

Lombardi, L. (1991). *Laryngeal features and laryngeal neutralization* (PhD. dissertation, University of Massachusetts, Amherst, MA. [Published by Garland, New York, 1994.]

Metsala, J. L. (1997). Spoken word recognition in reading-disabled children. *Journal of Educational Psychology*, 89, 159–69.

Ota, M., Yamane, N., & Mazuka, R. (2018) The effects of lexical pitch accent on infant word recognition in Japanese. *Frontiers in Psychology*, 8, e02354.

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 Disorders*, 9, 45–56.

R Development Core Team (2011). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

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

Shuai, L., & Gong, T. (2012). Voice onset time as a cue for perceiving place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 131, 3309.

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

Singh, L., & Chee, M. (2016). Rise and fall: effects of tone and intonation on spoken word recognition in early childhood. *Journal of Phonetics*, 55, 109–18.

Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, 124, 128–42.

Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: a cross-linguistic investigation. *Developmental Science*, 17, 94–109.

Singh, L., Poh, F. L. S., & Fu, C. S. L. (2016). Limits on monolingualism? A comparison of monolingual and bilingual infants' abilities to integrate lexical tone in novel word learning. *Frontiers in Psychology*, 7, e00667.

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

Son, G. (2017). *Interactive development of F0 as an acoustic cue for Korean stop contrast* (unpublished doctoral dissertation), University of Pennsylvania, Philadelphia, PA.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced–voiceless distinction for stops. *Journal of the Acoustical Society of America*, 55, 653–9.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B*, 364, 3617–32.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147–66.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form represenations of 14-month-olds. *Psychological Science*, 13, 480–4.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71, 73–108.

Trainor, L., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2), 394–402.

Tse, J. K. P. (1978). Tone acquisition in Cantonese: a longitudinal case study. *Journal of Child Language*, 5, 191–204.

Tuaycharoen, P. (1977). *The phonetic and phonological development of a Thai baby: from early communicative interaction to speech* (unpublished doctoral dissertation, University of London).

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.

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). The roles of voice onset time and F0 in stop consonant voicing perception: effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research*, 56(4), 1097–107.

Wright, J. D. (2007). *Laryngeal contrast in Seoul Korean* (unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.