INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

F0 development in acquiring Korean stop distinction

Gayeon Son

Department of Linguistics, University of Pennsylvania, USA

gson@ling.upenn.edu

Abstract

A number of studies have investigated the role of Voice Onset Time (VOT) on acquisition of stop voicing contrast. Korean stop contrasts (lenis, fortis, and aspirated), however, cannot be differentiated only by VOT since they are all pulmonic egressive voiceless stops. For this three-way distinction, another acoustic parameter, fundamental frequency (F0), critically operates. The present study explores how F0 is perceptually acquired and phonetically operates for Korean stop contrast over age. In order to reveal the relationship between F0 developmental patterns and age, a quantitative acoustic model dealt with word-initial stop productions by 58 Korean young children aged 20 months to 47 months. The results showed that phonetic accuracy depends on the perceptual thresholds in F0, and the significant phonetic differentiation with F0 between lenis and aspirated stops was significantly related to age. These findings suggest that acquisition of F0 plays a crucial role in the formation of phonemic categories for lenis and aspirated stops and this process significantly affects articulatory distinction.

Index Terms: F0 acquisition, quantitative acoustic analysis, age effect, Korean stop contrast

1. Introduction

Language learning begins with understanding speech sounds. Infants initially acquire human speech signals holistically, in meaningful chunks, but they soon realize that the chunks are sequences of sounds. During this process, they learn to recognize the phonetic parameters that make phonological contrasts in their native language and appropriately align speech signals to native phonetic boundaries. With these acquired phonetic parameters, they accept native contrasts but decline non-native contrasts, which ultimately allows them to interpret speech sounds [1]. Thus, the acquisition of native phonetic parameters and the discrimination of native contrasts are closely related. This process automatically occurs early in language development, and through assessing children's early meaningful speech, we can examine what phonetic parameters they have acquired and what native contrasts they can distinguish. The present study deals with the acquisition of a native phonetic parameter and its role in the development of distinction of native contrasts.

Stops are the Manner of Articulation that children acquire first cross-linguistically, and there is a universal acoustic parameter, Voice Onset Time (VOT), used in making stop contrasts across languages [2]. VOT represents the temporal relationship between the oral constriction release and the vocal folds vibration. There are effectively three stop phonation types that differ in VOT, particularly when a stop sound occurs utterance-initially: lead-voice (voicing), short lag (voiceless unaspirated), and long lag (voiceless aspirated). Thus VOT is a key phonetic parameter to investigate children's developmental patterns regarding stop distinctions and has been used to study the process of stop acquisition in relation to the articulatory achievements of young children (see e.g., [3]-[8]).

In spite of a number of studies on stop distinction in children, little research has reported on the acquisition of Korean stops. Korean has an unusual three-way contrast known as lenis, fortis, and aspirated, which are all pulmonic egressive voiceless stops. Since the three universal stop categories of lead, short lag, and long lag cannot account for the three-way distinction in Korean, VOT is not a useful tool in this case. Instead, fundamental frequency (F0) is a useful acoustic measure for the Korean stop distinction. F0 is generally used as a prosodic term representing pitch contour in a sentence. However, in Korean stop distinction, F0 at vowel onset can be critical to identifying a preceding stop, since F0 values differ depending on what stop category precedes a vowel. Because of this, it is considered an important phonetic parameter, even though it does not seem to be a universal one.

The present study broadly asks how young Korean children acquire the three-way stop distinction in relation to its acoustic parameter, F0. It is suggested that fortis stops are acquired first, followed by lenis and aspirated stops. In the acquisition of fortis stops, the role of VOT is determinant and is thus considered a universal mastery pattern [9]. However, the role of F0, which is a language-specific acoustic parameter, has not been discussed in terms of the acquisition of lenis and aspirated stops, in spite of the fact that without an F0 difference, the series of three stops cannot be phonetically differentiated at all. In addition, the perceptual capability of young children has not been previously studied, while children's articulation has been analyzed in the existing literature. Due to this discrepancy, it is still in question what phonetic categories children have at different developmental stages and how F0 affects the development of a distinction between lenis and aspirated stops. To answer these questions, the present project aims to investigate the perceptual capabilities of young Korean children and the interrelation between production and perceptual achievement with respect to stop distinction. Focusing on F0, the present study will show how articulatory distinction between lenis and aspirated stops depends on children's perceptual abilities. Another goal of the study is to provide large-scale normative data on children's articulation and to exhibit a pattern of development of a native contrast.

To pursue these goals, this study conducted experiments with 58 young children aged 20 to 47 months and provides a phonetic analysis of the data. The experiments consisted of a production test and a perception test. In the production test, to

http://dx.doi.org/10.21437/Interspeech.2016-651

elicit their natural production of Korean stops, a picture-naming task was used with near-minimal triplets. Finding appropriate minimal triplets that toddlers would know and that should be picture-describable would be almost impossible. In the perception test, a point-to-a-picture task was used with synthesized stimuli. Through these experiments, the phonetic accuracy of their stop production and the role of their perceptual ability in articulation accuracy can be analyzed.

2. Methods

2.1. Participants

2.1.1. Production

A total of 58 children varying in age from 1;8 to 4;0 (years;months) participated in the production experiment. Participants' gender was not considered for recruiting. All child participants were native Korean monolinguals without hearing or speaking disorders, as reported by their parents/guardians.

Table1: *Child participants' information for a production test.*

Age group	Male	Female
-2;0	2	2
2;1-2;6	5	6
2;7-3;0	5	8
3;1-3;6	8	8
3;7-4;0	3	11

2.1.2. Perception

Of the 58 children who participated in the production test, 48 (between 2;0 and 4;0) also participated in the perception experiment. Both experiments were conducted in Seoul, Korea.

Table 2: *Child participants' information for a perception test.*

Age group	Male	Female
2;0-2;6	5	6
2;7-3;0	4	4
3;1-3;6	8	7
3;7-4;0	3	11

2.2. Procedure

2.2.1. Production

A picture-naming task was used for the production experiment. Nine words, which included three types of stops (lenis /p, t, k/ - fortis /p*, t*, k*/ - aspirated /p^h, t^h, k^h/), were carefully selected, and each word was prepared as a describable picture. Each picture was shown to the participants three times in random order, restricted so that no picture followed itself successively. The words used were as follows.

Table 3: *Target words.*

	Lenis	Fortis	Aspirated
Labial	/paŋ/ 'room'	/p*aŋ/ 'bread'	/p^hal/ 'arm'
Alveolar	/tal/ 'moon'	/t*al/ 'daughter'	/t^hal/ 'mask'
Velar	/kɔŋ/ 'ball'	/k*ɔt/ 'flowers'	/k^hɔŋ/ 'beans'

An omni-directional YETI microphone was used for the recording, and it was placed at a distance of 3-15 cm from the participant's mouth. The whole session was recorded and digitized on Praat version 5.3.01 set up on a personal laptop at a 22050 Hz sampling rate and an 11025 Hz filter rate with 344 Hz of bandwidth. The recordings were conducted in a quiet room in a daycare center library, and saved as WAV files. Target words were extracted from the recordings and the word-initial stops were analyzed using Praat. Through this experiment, nearly 1,300 tokens were selected as appropriate for phonetic analysis. Mainly based on spectrograms, F0s at the following vowel onset was estimated as represented by the onset of the second formant.

2.2.2. Perception

A pair of pictures at a time was presented to each child participant using a personal laptop. The participant was asked to point to one of the two given pictures right after hearing "*mwueti* [target word] *ici?*" meaning 'which one is [target word]?' from the audio. The children were encouraged to touch the laptop screen, so if they touched one of the minimal pair pictures when they heard the stimulus, the pair of pictures for the next trial would pop up. The stimuli that were used in this point-to-a-picture task were synthesized stimuli. The stimuli were mixed in random order and provided to children in the test. The perception experiment was conducted directly after the production experiment finished in the same room at each daycare center. The session usually took 10 minutes. The session consisted of 17 trials (5 fillers + 12 synthesized stimuli). If a child seemed distracted, the same question was asked again. Every answer by the participant was documented by the experimenter during the test.

The productions of /pal/ ('foot') and /kɔŋ/ ('ball') by a female speaker were identically used for the synthesis. The onset of the following vowel was manipulated to have a 15 Hz difference between the tokens. Using Praat, the vowel onset of /pal/ was manipulated to have 200 Hz, 215 Hz, 230 Hz, 245 Hz, 260 Hz, and 275 Hz with fixed VOT (70 ms). The vowel onset of /kɔŋ/ has 220 Hz, 235 Hz, 250 Hz, 265 Hz, 280 Hz, and 295 Hz with fixed VOT (75 ms). For all cases, F0 contours were not changed and F1, F2, and F3 values were set to a fixed value so that resulting stimuli sounded natural.

Table 4: *Synthesized stimuli.*

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

3. Results

3.1. Production

Participant production patterns are illustrated in Figure 1. F0 values are quite widely dispersed from 100 Hz to 600 Hz across the three stop categories, but the median in each category shows some F0 differences: aspirated stops had the highest F0s, while lenis stops had the lowest F0s. At this point, it can be said that the child participants were not able to perfectly distinguish F0 between the fortis and aspirated stops, since the F0s for the two stop categories seem to be similar.

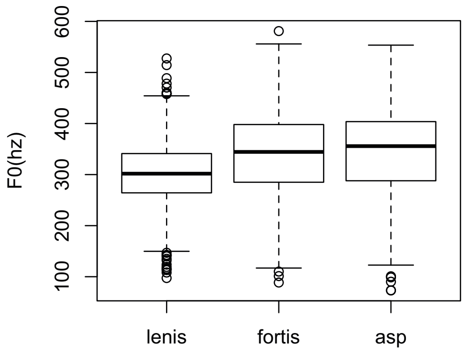


Figure 1: *Boxplot for F0 in three stop categories by children.*

To draw a relationship between age and the F0 values of the three stop series, a mixed effects linear regression model was designed. This multilevel model is considered most useful for this analysis because it is able to recognize that some sound tokens were produced by the same speaker if the same child speaker repeatedly produced the same target sound, and that the sound tokens include information from the two different levels, an individual and a production level. Using the mixed effects model, these two kinds of information can be distinguished and analyzed with separately. For this reason, this multi-level analysis was used to identify any correlation between F0 and age in the speech data. In this model, we use the lenis stops and female speakers as a baseline category, so including the interaction term, stop category (*Fortis* or *Asp*)**Age*, the equation (1) was used for the model.

$$F0_{ij} = \beta_0 + \beta_1 * Age_i + \beta_2 * Male_i + \beta_3 * Fortis_{ij} + \beta_4 * Asp_{ij} + \beta_5 * Fortis_{ij} * Age_i + \beta_6 * Asp_{ij} * Age_i + \eta_i + \varepsilon_{ij} \quad (1)$$

(i = individual level, j = production level)

Table 5: *The output of a mixed effects linear regression model.*

Random effects:

Groups	Name	Variance	Std. Dev.
child_id	(intercept)	935.6	30.6
Residual		5156.8	71.8

Number of obs: 1322, groups: child_id, 58

Fixed effects:

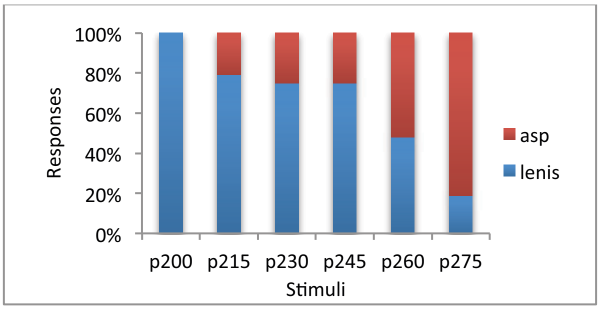
	Estimate	Std.Error	t value	Pr (>|t|)
(Intercept)	379.19	23.22	14.57	<2e-16 ***
Fortis	20.47	4.09	8.41	<2e-16 ***
Asp	-39.12	4.08	9.50	<2e-16 ***
Age	-2.06	0.69	-2.19	0.0699 .
Male	-27.46	9.10	-3.01	0.0072 **
Fortis:Age	0.35	0.60	0.61	0.53994
Asp:Age	2.10	0.59	3.35	0.00064 ***

Sig. codes: 0 '***' 0.001 '**' 0.01 '*' 0.5 '.' 0.1

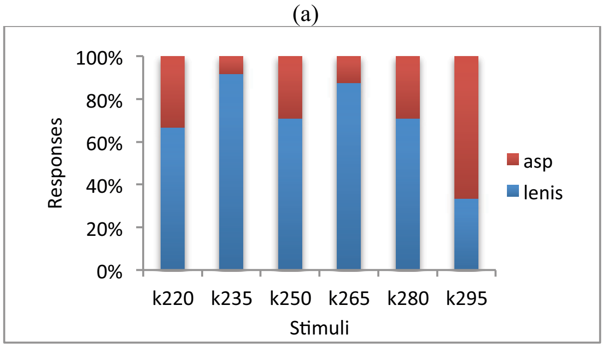
The results reveal that the effects of age can be found in the relationship with lenis ($p < .1$) and aspirated stops ($p < .001$), but not with fortis stops ($p = 0.53994$). According to the fixed effects coefficients, lenis stops tend to have lower F0s with age (coef. = -2.06) while aspirated stops tend to have higher F0s (coef. = 2.10) with significance ($p < .01$). That is, the phonetic gap, in terms of F0, becomes larger over age. This tendency implies that the acoustic differentiation between lenis and aspirated stops becomes more specific and more accurate as children get older. The insignificance of the relationship between age and fortis stops may indicate that F0 is not a critical acoustic parameter to phonetically (and phonologically) define fortis stops, which means that F0 accuracy would not contribute to the production of fortis stops even by older children. This possibility accords with the finding in [10] that only VOT plays a significantly important role in defining fortis stops. The significantly short VOT already functions enough to phonetically distinguish fortis stops, so another acoustic parameter F0 mainly serves to distinguish between the two stop categories, lenis and aspirated stops. Accordingly, there would be no need for distinctive F0 values for only fortis stops.

3.2. Perception

The results of the identification of the two contrastive pairs with slightly changed F0s at the following vowel onset reveal that perceptual F0 thresholds of the target children would be quite similar in the case of /p^h/. The following figures represent the ratio of all the responses for the two alternatives, a lenis and an aspirated stop.



(a)

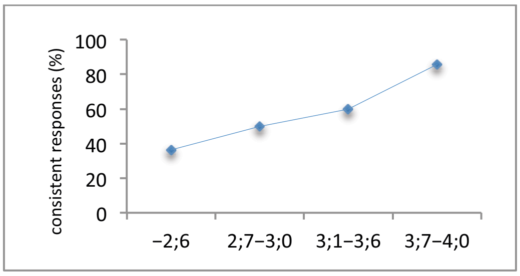


(b)

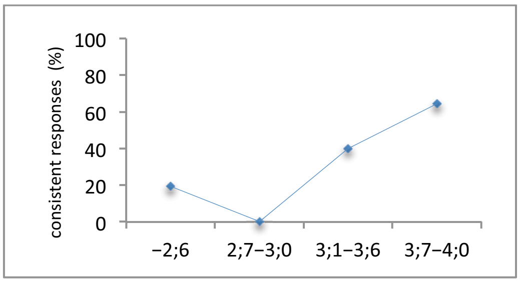
Figure 2: *Results of the identification of synthesized /p/ (a) and /k/ (b) by all the child participants.*

In the case of synthesized /k/, the children's labeling patterns were not unified or consistent. In spite of this fact, a 75 Hz difference from the standard lenis stop would be identified as its aspirated counterpart in both cases of /p/ and /k/. This finding would indicate that the child participants were able to recognize F0 differences as a key acoustic property of aspirated stops. The relatively confusing pattern in the identification of synthesized /k/ could be explained as the phonological difference in acquiring phonemes. The acquisition of bilabial stops would occur prior to that of velar stops, so this phonological ordering could affect the responses of the younger children.

The next step is to analyze children's responses in relation to their age. The analysis includes only valid responses, which are represented in the following figure.



(a)



(b)

Figure 3: *Consistent response levels in the identification test (synthesized /p/: (a), synthesized /k/: (b)).*

If a child labeled a given stimulus with a lower F0 as an aspirated stop, then other stimuli with higher F0s should be labeled as aspirated, too. If this condition was not met, those responses were excluded from the analysis. As expected, relatively older children showed more consistent identification than did younger age groups. However, in the case of synthesized /k/, the second youngest age group (2;7-3;0) showed no consistent labeling at all. This pattern indicates that until the age of 3, the consistent identification of the velar lenis or aspirated stops would be challenging for the children. The mean F0 differences to be required to label aspirated stops by the four different age groups are calculated as follows.

Table 6: *Mean F0 differences from lenis stops to be required for labeling aspirated stops by four age groups.*

Age groups	Mean F0 difference from lenis to asp (Hz)	
	/p/ to /p^h/	/k/ to /k^h/
-2;6	+ 45	+ 75
2;7-3;0	+ 60	N. A.
3;1-3;6	+ 60	+ 55
3;7-4;0	+ 65	+ 73

In both cases, perceptual F0 thresholds for aspirated stops are quite similar and the perceptual thresholds do not seem to be affected by age. How these F0 differences between lenis and aspirated stops in perception affect articulatory distinction by the same child participants remains in question. Therefore, for a comparison, F0 differences between the two stop categories were calculated from their actual productions, which were obtained through the production experiment (Table 7).

From the comparison of the F0 differences in children's actual perception and production, it would be suggested that their perceptual threshold in F0 would be responsible for their articulatory distinction between lenis and aspirated stops, since the older group almost reaches their perceptual F0 differences in their productions. For example, the youngest group was not able to make much F0 differences between /p^h/ and /p/ (16.4 hz difference), but the oldest group showed a quite enough distinction with 67.7 Hz between the two stop sounds. Thus, the age effect on articulatory and perceptual distinction is apparent. Younger groups did not successfully or consistently label the given stimuli, and F0 differences between lenis and aspirated stops are not enough to phonetically distinguish the two categories.

Table 7: *Actual F0 differences between aspirated and lenis stops in children's productions.*

Age groups	/p^h/-/p/ (Hz)	/k^h/-/k/ (Hz)
-2;6	16.4	49
2;7-3;0	22.2	N. A.
3;1-3;6	44.9	78.7
3;7-4;0	67.7	82

4. Discussion

The results indicate that in the acquisition of Korean stop contrast between aspirated and lenis stops, an acoustic parameter, F0, plays a key role. The articulatory distinction between the two different stop categories significantly depends on F0 values on the following vowel onset. These phonetic differences become apparent over age, which indicates that the development of F0 is in progress in acquiring Korean stop contrast. However, the acquisition of fortis stops does not seem to depend on F0 values, which has been also reported in [11]. The target aged children showed they have similar perceptual thresholds in F0 to phonetically determine aspirated stops, while their actual articulations of aspirated stops were quite different. This discrepancy between production and perception would indicate that perceptual learning occurs prior to production learning and that younger children are still acquiring their native acoustic parameter, F0.

5. Conclusions

This study explores the role of an acoustic parameter, F0, in distinguishing a language-specific stop contrast. It aims to investigate how Korean young children acquire and use F0 in their perception and production to make a phonetic distinction between lenis and aspirated stops through experiments. The results reveal that F0 development occurs over age and that this development is not completed before 2 years of age. The results also showed that articulatory distinction depends on perceptual discrimination.

6. References

[1] J. Werker and R. Tees, “Cross-language speech perception: Evidence for perceptual reorganization during the first year of life,” *Infant Behavior and Development*, vol. 7, pp. 49-63, 1984.

[2] L. Lisker and A. S. Abramson, “Cross-language study of voicing in initial stops: Acoustical measurements,” *Word*, vol. 20, pp. 384-422, 1964.

[3] D. Kewley-Port and M. S. Preston, “Early apical stop production: A voice onset time analysis,” *Journal of Phonetics*, vol. 2, pp. 195-210, 1974.

[4] M. A. Macken and D. Barton, “The acquisition of the voicing contrast in English: A study of Voice Onset Time in word-initial stop consonants,” *Journal of Child Language*, vol. 7, pp. 41-74, 1980.

[5] H. Clumek, D. Barton, M. A. Macken, and D. Huntington, “The aspiration contrast in Cantonese word-initial stops: data from children and adults,” *Journal of Chinese Linguistics*, vol. 9, pp. 210-224, 1981.

[6] G. Allen, “How the young French child avoids the pre-voicing problem for word-initial voiced stops,” *Journal of Child Language*, vol. 12, pp. 37-46, 1985.

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

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

[9] E. Kong, M. E. Beckman, and E. Jan, “Why are Korean tense stops acquired so early: The role of acoustic properties,” *Journal of Phonetics*, vol. 39, no. 2, pp. 196-211, 2011.

[10] E. Kong, *The development of phonation-type contrasts in plosives: cross-linguistic perspectives*. Ph.D. dissertation, Ohio State University. Columbus, Ohio, 2009.

[11] M. Kim, and C. Stoel-Gammon, “The acquisition of Korean word-initial stops,” *Journal of Acoustical Society of America*, vol. 125, no. 6, pp. 3950-3961, 2009.