

Durational Interaction of Stops and Vowels in English and Korean Child-Directed Speech

Choi, Hansook¹⁾

ABSTRACT

The current study observes the durational interaction of tautosyllabic consonants and vowels in the word-initial position of English and Korean child-directed speech (CDS). The effect of phonological laryngeal contrasts in stops on the following vowel duration, and the effect of the intrinsic vowel duration on the release duration of preceding stops in addition to the acoustic realization of the contrastive segments are explored in different prosodic contexts - phrase-initial/medial, focal accented/non-focused - in a marked speech style of CDS. A trade-off relationship between Voice Onset Time (VOT), as consonant release duration, and voicing phonation time, as vowel duration, reported from adult-to-adult speech, and patterns of durational variability are investigated in CDS of two languages with different linguistic rhythms, under systematically controlled prosodic contexts. Speech data were collected from four native English mothers and four native Korean mothers who were talking to their one-word staged infants. In addition to the acoustic measurements, the transformed delta measure is employed as a variability index of individual tokens. Results confirm the durational correlation between prevocalic consonants and following vowels. The interaction is revealed in a compensatory pattern such as longer VOTs followed by shorter vowel durations in both languages. An asymmetry is found in CV interaction in that the effect of consonant on vowel duration is greater than the VOT differences induced by the vowel. Prosodic effects are found such that the acoustic difference is enhanced between the contrastive segments under focal accent, supporting the paradigmatic strengthening effect. Positional variation, however, does not show any systematic effects on the variations of the measured acoustic quantities. Overall vowel duration and syllable duration are longer in English tokens but involve less variability across the prosodic variations. The constancy of syllable duration, therefore, is not found to be more strongly sustained in Korean CDS. The stylistic variation is discussed in relation to the listener under linguistic development in CDS.

Keywords: Voice Onset Time (VOT), vowel duration, prosody, focal accent, linguistic rhythm, English, Korean, Child-directed speech

1. INTRODUCTION

Duration of segments is conditioned by multiple factors. Some of the effects are from the neighboring contexts, and some other effects are from the inherent quality of individual segments. It has been reported that the vowel duration is intrinsically different based on the vowel qualities. Close vowels

are generally found to be shorter than open vowels (Naeser, 1970) and tense vowels are longer than lax vowels. Consonants also reveal durational variations based on segmental quality. In stops, different laryngeal features relate to different release durations, namely, voice onset time (VOT). VOT is widely accepted as a feature to mark the laryngeal contrast of stop consonants in most languages including English (Lisker & Abramson, 1964), such that English voiceless stops show greater VOT values than their voiced counterparts. Other laryngeal gestures as well as voicing are also reflected with differences in VOT. As for Korean stops, counted voiceless phonologically, manifestation of the three laryngeal categories is discussed

1) Haskins Laboratories, choi@haskins.yale.edu

mainly with the acoustic cues of VOT (Han & Weitzman, 1970; Choi, 2002; Jun, S.A., 1993; Kim, C., 1970; Kim, M., 2004; Lisker and Abramson, 1964 among others). Tense and aspirated stops are marked small VOTs and great VOTs respectively, compared with lax plain stops.

The duration interaction is intensively reported from vowel variation depending on the neighboring segments. A well-discussed effect is from the postvocalic consonant. It is generally agreed that the quality of the following consonant decides the duration of the preceding vowel. The vowel in English, for example, is longer when it is followed by a voiced consonant, than when it is followed by a voiceless consonant (House & Fairbanks 1953; Umeda 1975, for example). Peterson & Lehiste (1960) reported an effect of place of articulation of the following Cs in English, such that bilabial consonants condition reduced duration in the preceding vowels.

In contrast to numerous studies on the interaction between vowel duration and the postvocalic consonants, the research on the effect of prevocalic consonants is limited and findings are rather controversial. Naeser (1970) reported from English CVC syllables that vowel duration was increased significantly if the initial consonant was voiced rather than voiceless, which is also supported by French data (Thilly, 1993). The tendency of longer vowel duration after voiceless stops is observed in different speech styles such as a constant sentence frame (Peterson & Lehiste, 1960) and connected speech (Crystal & House, 1988). Ohala & Ohala (1992)'s study on Hindi reported different effects of voicing and aspiration on the following vowel duration such that voiced consonants are followed by shorter vowels, and aspirated ones have even shorter following vowels. The correlation between preceding VOT and following vowel duration is found nonlinear. Allen & Miller (1999) showed that the change from voiced to voiceless in English monosyllabic words resulted in decreased vowel duration as well as increased VOT and suggested "a trade-off" relation between VOT and vowel duration. But the decrease in vowel duration was consistently less than the increase in VOT, which is confirmed in two different speech rates. Similarly, Fischer-Jørgensen (1964) points out for Danish that the magnitude of the decrease in vowel duration following voiceless compared with voiced stops is insufficient to compensate fully for the increase in VOT. Thilly (1993) observed an effect of place of articulation on vowel duration in French such that vowels were longer after labial than after dental or palatoalveolar fricatives. Peterson & Lehiste (1960) and Strange et al. (1979), however, argued that

effects of prevocalic consonants on vowel duration were not very robust in American English.

To sum up, temporal reciprocation of tautosyllabic consonants and vowels has been observed in certain languages, but universality of the reciprocal pattern is not overtly confirmed across languages (Laver, 1994: 446). Language-specific variations are still found on the interaction, such as the diverging conclusions by Thilly (1993) versus Peterson & Lehiste (1960) and Strange et al. (1979). Furthermore, prosody as another plausible factor to induce effects on the tautosyllabic consonant-vowel interaction (de Jong, 1991) has not been investigated well.

The current study explores the quantitative interaction of tautosyllabic consonants and vowels with controlled prosodic context in another speech style, Motherese or child-directed speech (CDS), across languages of English and Korean. The style of Motherese is of specific interest here because several unique patterns in speech production are reported in CDS. In general, CDS is found to feature exaggerated prosodic realizations, such as exaggerated pitch peaks and extensive stress marking (Fernald and Mazzie, 1991; Fernald et al., 1989, for example), and more clarified consonant and vowel production (Kuhl et al., 1997; Malsheen, 1980; Ratner, 1984 for example). The current study investigates how the CDS features of the segmental clarification and prosodic exaggeration influence the durational realization of the stop and the vowel contrasts. English and Korean are compared here because they have different linguistic rhythms and different segmental contrast systems.

Korean stops have a 3-way laryngeal contrast and are counted voiceless phonologically unlike English 2-way voicing contrast. The English vowel system has a tense and lax distinction, whereas the Korean vowel system has a long and short distinction (Lee, H.B., 1999), though it is not very active in Modern Korean. The two languages have also different speech timing systems. English is generally counted as stress-timed whereas Korean is counted as syllable-timed or mora-timed (Abercrombie, 1967; Cho, M., 2004; Ladefoged, 1975 among others). The effect of prosodic context and linguistic rhythm on durational variation of VOT and vowel phonation is compared in English and Korean Motherese. It is also observed what the intrinsic effects are on the duration from the vowel and consonant contrast and how the consonant and vowel interact quantitatively.

Most studies consider the vowel duration as the vowel

phonation time that is visible after the consonant release, and in the articulatory viewpoint, the terminology is not exact. However, in the current study, the phonation time is measured as vowel duration consistently.

2. METHOD

2.1. Subjects

Speech data were collected from 4 Korean speaking mothers and 4 American English speaking mothers, who had 12 to 18 months old infants in the current study. The participants' target infants (E1, E2, E3, and E4 for the English setting, K1, K2, K3 and K4 for the Korean setting) were at the very initial period in the MLU (Mean length of Utterance) stage 1 and were producing very limited one-word utterances mostly at the time of observation. The American English speaking mothers who were native in the given language were all born in US and were staying in the New England area, US. Both of the parents were native speakers in American English and used English to their infants as the major language of communication. The Korean mothers were all born in Korea, lived in Korea until their puberty, and were native speakers of Korean though they were staying in US at the time of observation. The mothers and fathers, native in Korean, used Korean to their infants as the major language of communication. No hearing or speaking impairments were reported from both the mothers and their children. In the original recording, 10 mothers in total participated but acoustic analyses excluded results of one English speaking mother and one Korean speaking mother due to the inconsistent settings of recording and failure in interacting with the infants consistently. All of the subjects were paid for their participation.

2.2. Speech Material

English laryngeal contrast was observed from a consistent place of articulation, bilabial. Target consonants of /b/ and /p/ were analyzed from the initial CV syllable of six English words (with the target syllables underlined: potter, bottle, peter, beater, petter, bettor). The Korean stop contrast of /p/ 'plain stop', /p^h/ 'aspirated stop', and /p'/ 'tense stop' was observed from six existing Korean words (pata 'the sea', p^hata 'to dig', p'ata 'butter', pita 'to be empty', p^hita 'to bloom', p'ita 'to sprain'). All the target syllables for English tokens contained the lexical stress in the present study, and thus there was no variation of prosodic prominence at the lexical level. The words were

collected from different prosodic contexts, utterance-initial and -medial positions of minimally structured speech addressed to infants both in Korean and in English in a similar way. There was no prepared control on the sentence or focus. Target words were given as names for toys, and mothers had a playtime with their infants using the target toys. The conversation between the mother and the infant was recorded as a sound corpus. From the sound data, the target CV sequences were extracted, a total of 2199 tokens (1079 English + 1120 Korean), and identified for the prosodic condition individually. The boundaries of utterance were identified based on the obvious big pause and the context of the speech. Focus was identified based on the content of the speech as well as the acoustic patterns, such as pitch variation and loudness. The focus can be a contrastive narrow focus and a broad focus (Frascarelli, 2000; Ladd, 1996) with extra intensity and pitch-accent as well. Both types of focus are counted focus in the current study, which is denoted focal accent hereafter. Nonfocused tokens were accompanied with other obviously focused tokens within the same utterance.

2.3. Procedure

The participating mother interacted with her infant and shared conversation with the infant in a sound attenuated room at Haskins Laboratories, New Haven, CT, USA. The room was filled with toys, books, and a very comfortable couch in the middle to provide a comfortable setting to the mother and child. One Korean participant, K2, however, was observed at a silent room at the participant's house, with same toys and materials. The sound data didn't report any noticeable difference from the tokens recorded at Haskins Laboratories, and thus was included in the analysis.

The settings were identical in the English and Korean experiments except for the languages used. The target CV sequences were prepared as names of toys, and each toy was guided with a flash card containing its details. For example, 'Peter', a name of a stuffed animal, was printed in a bold letter and attached to the toy. In addition, flash cards were provided with information such as Peter's age, its favorite food and toys, best friends, and its family. Participating mothers were instructed to play with the children using the toys and the cards. They were asked to behave as usual as possible, and no explanation of the research goal and design was explained before the experiment. In order not to distract the infants, each toy was introduced separately as a new toy, and the mothers named the toy and described its characteristics using the

prepared lists of descriptions. After finishing one, a new toy was introduced replacing the old one. No sentences were prepared to be cited by the mothers in the introduction. The participants used their own sentences in their own styles and in their comfortable rates to give information about the toys.

At the end of the interaction, the mothers were asked to give lists of questions to the children. The questions were made in advance varying prosodic contexts and provided to Mothers in separate sheets. However, mothers, themselves, had to improvise answers to those questions or react to very few answers from their infants. This session was prepared to observe rather obvious patterns of focal accent variation. For example, for a question asking “is Peter a bear?”, the answer, “No, Peter is a cat”, provides a token of ‘Peter’ without the focal accent. On the other hand, in the question and answer of “is Peter a bear? No, Bettor is a bear”, ‘peter’ is a token with the focal accent.

The overall session lasted 40 to 50 minutes per individual mother and child. Optional long breaks were provided upon the subjects’ request. The interaction between mother and infant involved unrestricted short breaks frequently, and recording continued through the breaks.

The minimally structured speech of caregivers to their children was recorded using a wireless microphone and a TASCAM HD-P2 recorder at the sampling rate of 44100Hz. The recorded sounds were transferred to a PC and analyzed using Praat at the sampling rate of 44100Hz for the acoustic analysis. In total, 1079 CV sequences were collected from 4 American English speakers (E1: 117 tokens, E2: 127 tokens, E3: 431 tokens, E5: 404 tokens) and 1120 CV sequences from 4 Korean speaking mothers (E1: 337 tokens, K2: 202 tokens, K3: 170 tokens, K4: 411 tokens) as viable tokens after getting rid of invalid tokens such as noisy or devoiced tokens. In English, 385 tokens were from initial position, and 810 were with the focal accent. In Korean, 677 tokens were from initial position, and 818 tokens were with the focal accent. Each syllable has following number of tokens: English /ba/=160, /be/=181, /bi/ =184, /pa/=210, /pe/=176, /pi/=168; Korean /pa/=193, /p^ha/=176, /p^ha/=196, /p^hi/=191, /pi/=190, /p^hi/=174.

2.4. Measurement

The recorded sounds were transferred to a PC at a sampling rate of 44100Hz, and 1055 tokens of designated CV sequences were extracted and analyzed with the Praat program (Version 5.0.09, Boersma & Weenink, 2000). VOT and the phonation time after the stop release were employed as acoustic correlates

of the stop quantity and the vowel quantity respectively.

The duration from the stop release to the onset of the voicing in the following vowel was measured as VOT. The duration from the first peak of the stop release burst up to the zero crossing nearest to the onset of the second formant of the following vowel is measured based on the wave forms and the corresponding spectrogram. The vowel phonation period is measure from the onset of the second formant and the offset of the second formant of the spectrogram. The syllable duration is measured from the time point of stop release to the cessation point of the second formant of the vowel portion.

To calculate the variability in the production of segments and syllables, the measured acoustic values for each token were transferred into an edited version of Delta values which indicate the distance from the mean (McMillan & Corley, 2010 for example). The formula is used for the acoustic value of X:

$$\text{Delta } X_i = | X_i - \bar{X} |$$

(where \bar{X} = mean of X, X_i =individual measurement)

The mean VOT and vowel duration were calculated within individual speakers separately for each syllable type (/pa/, /ba/, /pi/, /bi/, /pe/, /be/ in English and /p^ha/, /pa/, /p^ha/, /p^hi/, /pi/, /p^hi/ in Korean), and Delta values were taken as absolute values of difference between the means and the individual acoustic values. For example, Delta_C is measured for ‘pi’ by subtracting the measured VOT from the mean VOT of ‘pi’ tokens. The current delta calculation measures variability within a category and excludes variations between different segments or segmental sequences. The bias from inherent features between different segments is excluded in the delta measure.

2.5. Statistical Analysis

The Welch T-test was employed to compare the overall difference between the two language groups, English vs. Korean. The influence of each prosodic factor and individual variation was evaluated based on Linear Mixed models. Analyses of Linear Mixed models were selected to detect overall significant effects across less balanced sets of tokens based on the flexibility of the model (Hoffman & Rovine, 2007). The tokens in this study were selected from unstructured conversational settings without intended repetition and thus, the number of tokens is not evenly distributed for different conditions as in repeated read speech experiments. Linear Mixed models were performed several times to measure different

factors taking the acoustic measurements as dependant variables. The initial Linear Mixed models selected speakers as random factors, and Position (i.e., Initial, Medial), Focus (i.e., Focused, Nonfocused), and Token types (i.e., (/pa/, /ba/, /pi/, /bi/ in English and /p^ha/, /pa/, /p^ha/, /p^hi/, /pi/, /p^hi/ in Korean) as fixed factors to measure overall effects of prosody and segments. Based on the result that shows the insignificant effect from the position, other factors were employed as fixed factors which is discussed in the result section in detail.

The SPSS statistical package (SPSS for windows, Standard Version, Release 11.0.1, 15 Nov 2001, SPSS Inc.) and RStudio (Version 0.95.262) were used as analyzing tools. In addition to statistical comparisons, distributions of individual subjects' tokens in each condition were compared using diverse charts and graphs.

3. RESULTS AND DISCUSSION

Statistical analyses revealed difference between the two language groups. Due to the sample distribution, Welch Two Sample t-test compared the difference between the two language groups. The summary is introduced in Table 1.

Tabel 1. Summary of Welch Two Sample t-test

	English Mean	Korean Mean	t-value	p-value
VOT	0.0479	0.0475	0.186	0.8528
Vowel duration	0.1337	0.1024	9.739	< 2.2e-16
Syllable duration	0.1816	0.1499	9.295	< 2.2e-16
Delta_VOT	0.01528	0.0143	1.201	0.23
Delta_Vduration	0.0325	0.0407	-3.478	0.000519
Delta_Sduration	0.04114	0.0801	-14.785	< 2.2e-16

The overall VOT values do not show difference between groups whereas the vowel duration shows difference which leads to the difference in syllable duration consequently. Table 1 presents very little difference of mean VOT and mean Delta_VOT values between the English data and the Korean data, and the difference is not statistically significant. Vowel duration and syllable duration show the significantly greater value of the English tokens in Table 1. However, the significant differences in the Delta Vowel duration and Delta syllable duration are due to the significantly greater value in the Korean data. In other words, the Korean Motherese features shorter

vowel duration with greater variability.

It is true that the laryngeal contrast systems in the two languages have difference and thus, more detailed comparison should be done. Linear mixed model analyses using Focus (focused vs. nonfocused), Position (initial vs. medial), and Token Types (4 English syllables + 6 Korean syllables) as fixed factors and Speaker (10 speakers) as a random factor reported significant main effects of Focus ($F(1,9.810)=16.685, p<.01$), Token Type ($F(10,32.373)=28.136, p<.001$), and Speaker ($F(6,21.205)=2.608, p<.05$). The focused tokens are greater than the nonfocused ones in VOT values. There is an interactive effect between Focus and Token Types ($F(10, 43.630)=3.387, p<.01$). The focus-induced durational increase is obvious in the English voiceless and Korean aspirated and plain tokens, as is depicted in Figure 1. However, the VOT increase is very minor in English voiced tokens and the Korean tense vowels even display decrease under focal accent. In the overall VOT values, the Tukey HSD posthoc comparison shows that English voiceless and Korean aspirated are significantly higher than the other token types. Korean plain tokens still have greater VOT values than English voiced and Korean tense tokens. The mixed model analyses were repeated dividing the factor, Token Type, as two separate fixed factors of Stop Laryngeal Contrast (voiced, voiceless, tense, aspirate, plain) and Vowel Contrast (English /a/, /i/, /ε/ and Korean /a/, /i/) as fixed factors and the effect of Stop Laryngeal Contrast was found significant ($F(3,9.465)=63.822, p<.001$) whereas Vowel Contrast was not significant ($F(3,9.465)=2.122, p=.164$). However multiple comparisons using Tukey HSD showed that English /a/ and Korean /a/ vowels correspond to significantly longer VOT values of the preceding stops at the 0.05 level. More detailed distribution of VOTs is observed in Figure 1, which displays average VOT values for the English and the Korean tokens divided by the dashed line and shaded in a different degree.

The t-test reported that mean distributions of the selected vowels in English and Korean motherese are not identical. The tested English vowels in the given setting were produced longer than the test Korean vowels in general. Even when the t-test compared the language group distinctions in the subsets of the token group with /a/ and the token group with /i/ in English and Korean, which are rather matching vowel qualities in the two languages, the results still show significant differences between groups with longer durations in English (/a/: $t = 7.7119, p = 3.175e-14$; /i/: $t = 14.049, df = 708.996, p < 2.2e-16$). The linear mixed model analyses using Focus, Position

and Token type as fixed factors and Speaker as a random factor reported significant variations due to Focal context ($F(1,7.463) = 22.548, p < .01$) and Token Types ($F(10,44.931) = 21.250, p < .001$). The focused vowels were longer than the nonfocused counterparts. The tokens with /a/ are longer in general among the token types except for the Korean aspirated /p^ha/ that shows a homogeneous subset with Korean /pi/, /pⁱi/, and English /pe/, /pi/ in Tukey HSD posthoc comparison at the alpha level of 0.05.

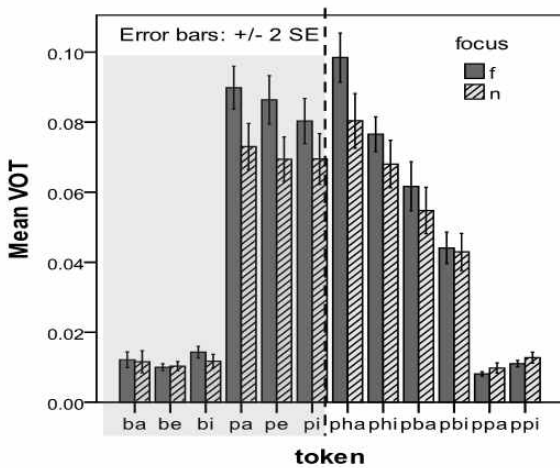


Figure 1. Average VOT for tokens in English (the left of the line, marked the voiced as bi /bi/, be /be/, ba /ba/ and the voiceless pi /pi/, pe /pe/, pa /pa/) and in Korean (the right of the line, marked the aspirated: phi /p^hi/, pha /p^ha/; the plain: pbi /pi/, pha /pa/; and the tense: ppi /pⁱi/, ppa /p^aa/)

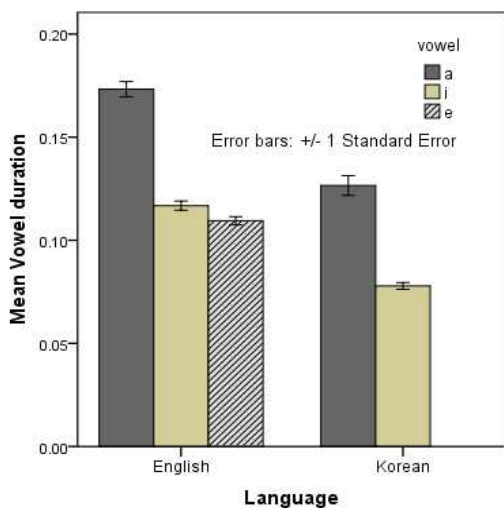


Figure 2. Average duration across subjects and preceding consonants (**p<.0001, *p<.05, 'e' is for /e/)

The durational difference is even more obvious in Figure 2. The vowel quality makes clear durational difference in English and Korean. In a linear mixed model with fixed factors of Vowel Contrast, Consonant Contrast, and Focus and a random factor of Speaker, the effect of Vowel Contrast was found significant ($F(3,14.989)=58.496, p < .001$). Detailed comparison using Welch t-test reports /a/ vowel tokens have significantly longer durations the non-low vowels while /i/ is still slightly longer than /e/. The factor of Consonant Contrast showed a significant effect in the vowel duration changes ($F(3,12.828) = 37.205, p < .001$). In contrast to the little effect of Vowel Contrast in the preceding VOT values, the Consonant Laryngeal Contrast shows a very robust effect on the following vowel phonation period, confirming previous studies on the asymmetrical interaction between consonants and vowels (Allen & Miller, 1999; Fischer-Jørgensen, 1964). Tukey HSD multiple comparison (at alpha=0.05) reports a significant longer vowel duration after English /b/, and a homogeneous group of English voiceless, Korean plain and tense. The Korean aspirated tokens showed the significantly short duration of the following vowel. In other word, the trade-off relationship between stop release duration and the vowel phonation time is detected in the current study such that English voiced tokens with significantly small VOTs precede vowels with significantly longer duration and Korean voiceless tokens with significantly great VOTs precedes vowels with significantly shorter durations. Pearson correlation measures also support the trade-off with a significant coefficient value (-0.168, Sig.<.001 at the 0.01 level, 2-tailed). The interactive patterns are clearly detected in Figure 3 consistently across the different vowel qualities.

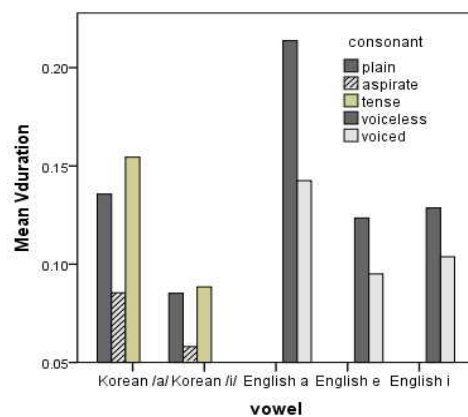


Figure 3. Mean duration of different vowels in English and Korean labeled in terms of preceding stop laryngeal contrast (English e denotes the vowel, /e/).

The focal effects display further complication due to the interaction with Token Type ($F(10,50.184)=2.343, p<.05$). The durational increase is greater in English /ba/ and Korean /pa/ and /p'a/. In other words, the open vowel involves greater durational increase compared to the other vowel qualities and the vowel after the stop without aspiration, that is marked with relatively smaller VOTs, shows the robust increase due to focal prominence. It is contrary to the increase patterns of VOT under focal accent in that the tokens with aspiration displayed greater increase under focal accent. However, the patterns can be understood as complementary trade-off between aspiration and voicing. The huge increase in the VOT of the tokens with aspiration leads to the limited increase in the vowel duration and vice versa. The negative correlation is also supported statistically by the negative correlation coefficient value of Spearman rho, $-.318 (p<.001)$. The interaction between vowels and consonants is specified in Figure 4. The focused tokens marked with bars of the plain pattern are greater than the nonfocused counterparts with the dashed pattern. The VOT variation depicted in the front light bars shows more robust increase in the English voiceless tokens (pi, pe, and pa in the figure) and the Korean aspirate tokens (phi and pha in the figure), whereas the vowel duration in the back dark bars shows more robust increase in the open vowel marked with 'a' in the figure. The paradigmatic increase dealing with individual contrastive segments leads to greater difference of the acoustic targets, which will be benefit to the infants learning the language eventually. However, the increase of individual segments shows an interaction with the neighboring segments, a trade-off relationship marked with the negative correlation between the quantities. In Figure 4, the trade-off relationship is captured from examples with /a/ vowels in English and Korean. The amount of increase due to the focal accent is limited in both languages when the preceding stops have great VOT increases like 'pa' in English and 'p'a' in Korean. The overall syllable durations of English /pa/ and /ba/, as a consequence of the trade-off, are produced very similar in each prosodic context.

However, the asymmetry in the interactive effects is also found such that the vowel changes are greatly affected by the preceding VOT changes, but the VOT values do not involve great influence from the following vowel duration. English VOTs are consistent across diverse vowel contexts, or Korean aspirate and plain tokens tend to have greater VOTs before 'a' than 'i'. As a consequence of the asymmetric interaction of

VOTs and vowel phonation time, the overall syllable duration is more dependent on the vowel quantity rather than the stop contrast. The intrinsic durational difference of vowels is not compromised by the VOT, in contrast to the VOT difference compensated by vowel duration, and thus decides the syllable duration. In Figure 4, constant duration of syllables is found among tokens with same vowels rather than same consonants.

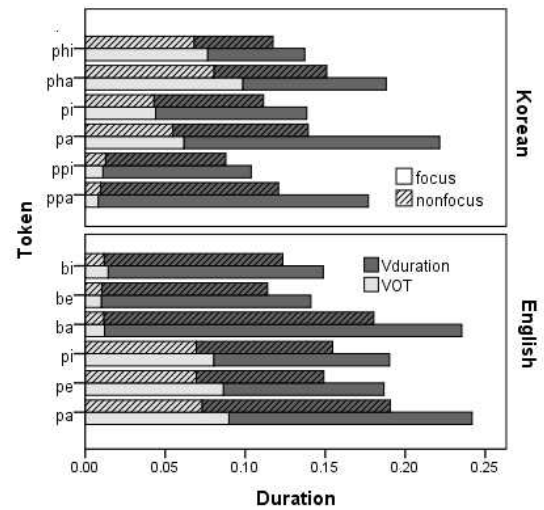


Figure 4. Average syllable duration as combination of VOT (light side) and Vowel duration (dark side) under two different focal contexts in English CDS (voiced: bi/bi/, be /be/, ba /ba/; voiceless pi /pi/, pe /pe/, pa/pa/) and in Korean CDS (aspirated: phi /p^hi/, pha/p^ha/; plain: pbi /pi/, pba /pa/; tense: ppi /pⁱi/, ppa /p^aa/)

Finally, the effect of linguistic rhythm was observed through comparisons of the variability of the syllable duration in English and Korean Motherese. Figure 5 presents mean Delta_syllable values, the difference from the average syllable duration of each token. The Delta calculation was conducted only within the group of same CV sequences to exclude any extra variation between the segments selected for the study. In the Delta measure, therefore, the variation among different segments such as durational difference between tense and lax vowels or voiced and voiceless consonants is excluded.

Figure 5 shows that the variability of syllable duration tended greater when the syllables are produced with the focal accent across speakers of English and Korean. All the subjects except for E2 show increased variability of the syllable with focal accent. The E2 is unique in that the nonfocused targets show greater variability that also shows a wide variation. If the exceptional case is not considered, the current results indicate that the observed CDS features exaggerated acoustic features

under focal accent and that the exaggerated targets involve more inconsistent realization of the durational targets. It is also noticed that the language difference still exists in the syllable variability. The Korean mothers overall show significantly greater variability in syllable durations than the English mothers both under focal accent and without focal accent ($F(1,2198)=213.801, p<.0001$). It is quite interesting since this result is not supporting the timing pattern in the linguistic rhythm. The Korean language with a syllable-timing is expected to maintain isochronous syllable durations more consistently compared to English, the typical stress-timed language. However, current results on CDS indicate that the overall production of syllables involves smaller durational variation in English though they are longer than the syllables in Korean CDS. The results are also contrary to the findings from a study on English and Korean adult-directed speech using same conditions and materials (Choi, 2012). In the ADS study, the linguistic rhythm was found to correspond to the variability difference such that Korean syllables were produced with less variability than English syllables across various contexts and the correlation between VOT and vowel phonation period was greater in Korean syllables.

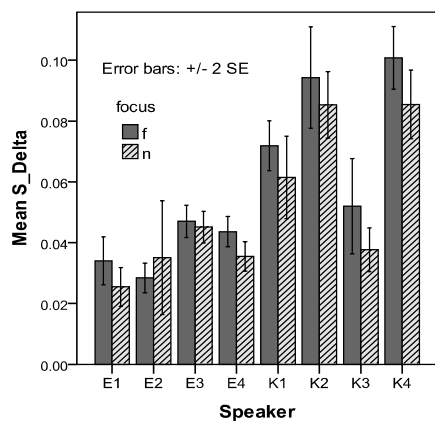


Figure 5. Mean of the variability measure of English syllables and Korean syllables for individual speakers under focused (f) and nonfocused (n) contexts.

4. CONCLUSION

This study explored the acoustic interaction between stop VOT and vowel phonation period within a syllable in Korean and English CDS under prosodic variation. The acoustic cues demarcate contrasts in consonants and vowels of both languages. The three-way and two-way laryngeal contrasts were specified

with different VOT values that become more distinctive under focal accent. Similar VOT resolutions were found in Korean tense stops and English voiced stops and in Korean aspirate and English voiceless stops. Vowel duration also shows intrinsic variation in that the open vowels are systematically longer. English vowels in the current setting were produced longer than Korean vowels. Focal effects are consistently found in both languages. The difference between the contrastive segments is enhanced under focal accent, supporting the paradigmatic strengthening effect. The contrastive segments with greater acoustic values, namely aspirate and voiceless tokens with longer VOTs and the open vowels with longer duration, show robust enhancement due to the focal accent. The increases of the segments with smaller acoustic values, on the hand, are relatively trivial. The positional difference was not found significant in the current setting. The tautosyllabic consonants and vowels also reveal an interactive pattern in English and Korean, which is asymmetrical. The durational variation of vowels is affected by the amount of the preceding VOTs, but not vice versa. Therefore, the acoustically realized syllable duration from the stop release to the vowel phonation tends to follow the intrinsic durational variation of vowel. The overt difference in contrastive consonants and vowels in terms of VOTs and vowel phonation period is not different from the prediction on CDS, of which the key feature is suggested as more clarified consonant and vowel production. This study further suggests that the enhancement of segmental contrast leads to different syllable durations for different CV types. The unit of syllable does not keep a constant size as a result of marking the segmental differences. How the infant takes quantitative differences of the same unit is the next question to be explored. It may suggest that mothers speaking to infants who are producing one word utterances are likely to focus on the segmental contrast rather than the global structures or units. Another result supports this idea. In this study, overall vowel duration and syllable duration are longer in English tokens but involve less variability across the prosodic variations. This result is different from the patterns in the speech directed to adults. The constancy in syllable durations is expected to be more faithfully observed in syllable-timed languages, here in Korean, and ADS showed the expected variation. The expectation, on the other hand, is not supported by the results of Motherese in English and Korean. The difference is taken as a uniqueness of the speech style, directed to word learning infants. The listener-tuned speech style seems to prioritize the segmental

contrast over the global rhythmic structures of languages. If so, the speech directed children at different stages of language acquisition is predicted to feature different temporal resolutions and prosodic patterns. Further study is therefore suggested to compare ADS and CDS in a more controlled setting.

REFERENCES

- Abercrombie, D. (1967). *Elements of general phonetics*, Columbia University Press.
- Allen, J. & Miller, J. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words, *Journal of the Acoustical Society of America*, 106, 2031-2039.
- Cho, M. (2004). Rhythm typology of Korean speech, *Cognitive Process*, 5, 249-253.
- Choi, H. (2002). Acoustic cues for the Korean stop contrast - dialectal variation, *ZAS Papers in Linguistics*, 28, 1-12.
- Choi, H. (2012). The Durational Interaction of Vowels and Stops under Prosodic Variation: A Cross-Language Study on English and Korean, manuscript.
- Crystal, T. & House, A. (1988). Segmental durations in connected-speech signals: Current results, *Journal of the Acoustical Society of America*, 83, 1553-1573.
- de Jong, K. (1991). An Articulatory Study of Consonant-Induced Vowel Duration Changes in English, *Phonetica*, 48, 1-17.
- Fernald, A. & Mazzei, C. (1991). Prosody and focus in speech to infants and adults, *Developmental Psychology*, 27(2), 209-221.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mother's and father's speech to preverbal infants, *Journal of Child Language*, 16, 477-501.
- Fischer-Jørgensen, E. (1964). Sound duration and place of articulation, *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung*, 17, 175-207.
- Frascarelli, M. (2000). *The syntax-phonology interface in focus and topic construction in Italian*, Dordrecht: Kluwer.
- Han, M. & Weitzman, R. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /ph, th, kh/, *Phonetica*, 22, 112-28.
- Hoffman, L. & Rovine, J. (2007). Multilevel models for the experimental psychologist: Foundation and illustrative examples, *Behavior Research Method*, 39, 101-117.
- House, A & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels, *Journal of the Acoustical Society of America*, 25, 105-113.
- Jun, S. (1993). The Phonetics and Phonology of Korean Prosody, PhD dissertation, Ohio State University.
- Kim, C. (1965). On the anatomy of the tensivity feature in stop classification: with special reference to Korean stops, *Word* 21, 339-359.
- Kim, M. (2004). Correlation between VOT and F0 in the Perception of Korean Stops and Affricates, *Proceedings of the 8th International Conference on Spoken Language Processing (ISCA)*.
- Kuhl, P., Andruski, J., Chistovich, I., Chistovich L., Kozhevnikova, E., Ryskina, V., Stolyarov E., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants, *Science*, 277, 684-686.
- Ladd, R. (1996). Intonational phonology, *Cambridge Studies in Linguistics*, 79, Cambridge: Cambridge University Press.
- Ladefoged, P. (1975). *A course in phonetics*, New York: Harcourt Brace Jovanovich.
- Lee, H. B. (1999). Korean, *Handbook of the International Phonetic Association*, Cambridge University Press, 120-122.
- Lisker, L. & Abramson, A. (1964). A cross-linguistic study of voicing in initial stops: acoustic measurements, *Word*, 20, 384-422.
- Naeser, M. (1970). Influence of initial and final consonants on vowel duration in CVC syllables, Technical Report No. 130. *Wisconsin Research and Development Center for Cognitive learning*, University of Wisconsin, Madison.
- Malsheen, B. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children, *Child phonology*, Vol. 2: Perception.
- McMillan, C. & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation, *Cognition*, 117, 243-260.
- Ohala, M. & Ohala, J. (1992). Phonetic universals and Hindi segment duration, In Ohala, J. J., Nearey, T., Derwing, B., Hodge, M., & Wiebe, G. (eds.) *Proceedings, International Conference on Spoken Language Processing*, 831-834.
- Peterson, B.; Lehiste, I. (1960). Duration of syllable nuclei in English, *Journal of the Acoustical Society of America*, 32, 693-703.
- Ratner, B. (1984). Patterns of vowel modification in mother-child speech, *Journal of Child Language*, 11, 557-578.
- Strange, W., Edman, T., & Jenkins, J. (1979). Acoustic and phonological factors in vowel identification, *Journal of*

Experimental Psychology: Human Perception and Performance,
5, 643-656.

Thilly, C. (1993). Influence of prevocalic consonant on vowel duration in French CV[p] utterances, *EUROSPEECH*, 93, 465-468.

Umeda, N. (1975). Vowel duration in American English, *Journal of the Acoustical Society of America*, 58, 434-445.

• **Hansook Choi**

Research Affiliate (2009~present)

Haskins Laboratories

New Haven, CT 06511 USA

Email: choi@haskins.yale.edu