VOWEL DURATION AS A PERCEPTUAL CUE FOR PRECEDING STOP LARYNGEAL CONTRAST IN KOREAN

Hansook Choi

Haskins Laboratories, USA
choi@haskins.yale.edu

ABSTRACT

This study explores the effects of vowel duration in identifying the previous stop laryngeal contrast in Korean. The durational resolution in Korean shows a limited variation of syllable duration and a strong correlation between VOT and vowel duration [1]. The systematic variation of vowel duration due to different VOTs of prevocalic stops suggests vowel duration as a plausible perceptual cue in identifying the preceding stops. Identification tests on stop categories (plain, tense, aspirated) in Korean were conducted with 12 native Korean listeners, using 33 disyllabic word tokens in the utterance-initial position. All sound tokens were manipulated in the vowel duration, pitch, and VOT. The results show that listeners employ multiple cues of VOT, pitch, and vowel duration in identifying the stop categories. It is also confirmed that vowel duration is actively used, even with unambiguous F0. The effect of vowel duration, however, is found to depend on stop laryngeal categories. Identification of the plain stop based strongly on F0 information. The effects of vowel duration were more obvious in tense and aspirated stops in the higher F0 ranges.

Keywords: VOT, vowel duration, F0, perceptual cue, laryngeal contrast, Korean stops

1. INTRODUCTION

The voice onset time (VOT) is widely accepted as a primary feature to signal the voicing contrast of stop consonants in most languages including English [5], such that English voiceless stops show greater VOT values than their voiced counterparts. Other laryngeal gestures besides voicing are also reflected in the different VOT ranges. In Korean with three voiceless stop laryngeal contrasts, for example, VOT still distinguish the three categories. The tense and the aspirated stops in Korean are marked very small VOTs and very great VOTs respectively, compared to plain unaspirated stops in the mid ranges ([3, 5] among others).

Another well-known acoustic cue for the laryngeal contrast is fundamental frequency (F0) at the following vowels. Whalen et al. [7] reports that F0 at the vowel onset assists voicing distinction of the preceding stops even with unambiguous VOT values. Kim, et al. [4] also suggests a very significant role of F0 of the following vowel portion in cueing the preceding stop consonants in Korean. The F0 perturbation in Korean due to the preceding stop contrast is suggested to extend to the mid point of vowel portion, which is discussed in relation to tonogenesis.

Other quantitative cues for the stops are further discussed from the neighboring segments, based on durational variations of vowels conditioned by following stops. The vowel is longer when it is followed by a voiced consonant than when it is followed by a voiceless consonant, and this consistent difference signals the voicing contrast of postvocalic stops, [2, 6] for example.

In contrast to numerous studies on vowels and postvocalic consonants, the research on the prevocalic consonant-to-vowel interaction is limited. Recently, the durational interaction is discussed based on language-specific temporal resolutions such as stress-timed and syllable-timed, and the compensatory quantitative interaction is more strongly confirmed in the syllable-timed language, namely Korean [1]. Korean is reported with small variability of overall syllable duration in the CVCV structure across varied prosodic contexts and speakers. The vowel shows quantitative variation due to the preceding stop's laryngeal feature, such that aspirated stops with great positive VOT corresponds to small vocalic period whereas tense stops with very small positive VOT corresponds to great vocalic period. The consistent durational difference in the vowel conditioned by the preceding stop laryngeal contrast suggests a plausibility of vowel duration as an acoustic cue for prevocalic stops, as is found in the vowel-to-postvocalic stop interaction. The current study is, therefore, to examine the perceptual reactions to the durational variation of

vowels in identifying preceding stops. It is explored what the significant role of vowel duration is in perceiving the laryngeal contrast of preceding stops, and how the durational cues interact with F0, another significant acoustic cues for the laryngeal contrast.

2. METHODS

2.1. Material

33 sound tokens were used as listening targets. The tokens were from one male native Seoul Korean speaker's read speech, saying *'pata / p^{h}ata / p'ata nɨn kɨ yeki-ɨi hekʃim-i-ya'* with the target underlined, meaning 'the sea/digging/butter is the key point of the story.' The male speaker produced 72 tokens with the given contrast in the utterance-initial position – the aspirated (*p^{h}ata*), the tense (*p'ata*), and the plain (*pata*) –, and average values of F0 and vowel duration were calculated. Vowel duration was measured from the beginning to the cessation of the second formant, and F0 was measured in the mid point of vowel using Praat (version 5.1. 23). Table 1 shows the calculated average values. Mean vowel duration categorize the tense as the longest and the aspirated as the shortest, whereas mean F0 categorizes the plain as the lowest and the aspirated as the highest.

Table 1: Average Vowel duration and F0 of speech tokens.

Laryngeal types	V duration	F0
Plain	0.07838	123.564
Aspirated	0.05939	160.281
Tense	0.09477	149.646

Based on the average values, the target syllables of the stimuli were modified with respect to VOT, F0, and vowel duration. To get rid of the effects of the well-known acoustic cue, VOT, the period from the stop release to the onset of voicing was sliced out at the zero-crossing point of the waveform. Vowel duration of each token was either extended or reduced in order to produce the three average values in Table 1. For example, the long vocalic period in the tense one, '*p'ata*', was sliced out to contain V durations of 0.078 second (sec.) and 0.059 sec. respectively. The overall deviation was within +/– 0.001 sec. from the mean values, and the editing produced 3 tokens of one F0 but with different V durations. Conversely, the aspirated token was extended by repeating cycles in the middle, so that longer vowels matching to the other two contrastive types were produced. Pitch is edited through the pitch edit function in Praat, and a token varied into the three mean F0 values. Additionally, 6 tokens with unedited VOT values were included in the stimuli, including three original tokens, to see effects of VOT. All in all, 33 tokens were used as sound stimuli (27 tokens with zero VOTs (3-way contrast x 3 pitch values x 3 vowel duration) + 6 tokens with VOTs).

2.2. Subjects

12 Korean native speakers (5 Females and 7 males) participated in the current study. The subjects [ages 33~43] were all born and educated in Korea up to the college or above levels. No hearing or speaking impairment was reported from all the participants. They voluntarily participated in the given study.

2.3. Procedure

The subjects listened to edited sound tokens in a quiet place through a headphone and identified the played sound among the options of *p^{h}ata, p'ata,* and *pata*. All the options and instructions were printed in Korean in the answer sheets. Additional oral instructions and practice tests were conducted prior to the real test. The subjects were told to listen to sound tokens, introduced as noised sounds with poor sound quality, and to provide their best answers for given tokens. Subjects were instructed to report the confidence of each answer after choosing the answer. Confidence levels were marked as 4 gradual categories (0%: don't know at all, 30%, 60%, 90%: very confident). Sound tokens were played for three times, and subjects selected the answer after listening to all the repetitions.

2.4. Analysis

The answers were summarized in terms of raw frequency and overall percentage within answers for individual questions. To measure significant effects from acoustic cues, the bivariate Pearson correlation analyses and univariate analyses of variance (ANOVA) were performed using PASW Statistics 18.

3. RESULTS

Pearson correlation coefficient measures report significant correlations between the most frequent answers and the discussed acoustic cues. The correlation was measured between the types of acoustic cues and major identifications. Significant correlations were between the identified tokens

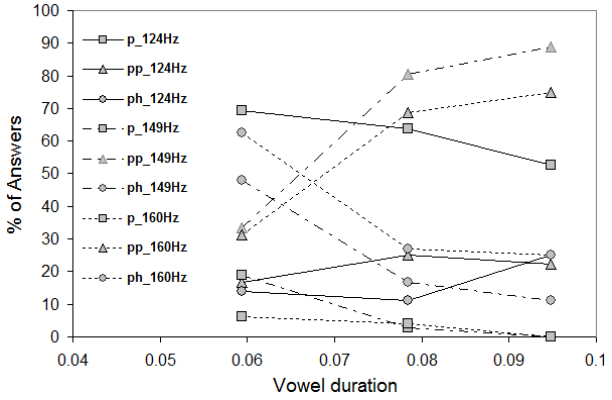
and VOT (r =.403, p<.05), F0 (r =.765, p<.01), and vowel duration (r =.348, p<.05). There is no significant overall correlation between original tokes and the identified tokens. The correlation indicates that the subjects identified the sound stimuli based on multiple acoustic values, and that the original sound type is not faithfully identified after the given acoustic cues are edited, suggesting weaker influence of the other vocalic features in prevocalic stop identification. Vowel duration, on the other hand, is found as a perceptual cue for prevocalic stop identification in Korean.

VOT is found as a significant cue in the present study, but not the only one. The strong correlation indicates very frequent answers for the tense after hearing the tokens with zero VOT. It should be still noted that the other stop categories were also identified for the sounds with the zero VOT. The answers even include identifications of the tokens with zero VOT as the aspirated. Seven questions with zero VOT were answered the aspirated as the most frequent answers. The results clearly support that F0 and vowel duration further signals the aspirated tokens.

More detailed effects of each acoustic cue on individual stop identification are reported from Univariate ANOVA on the proportion of each stop token, taking *Original type*, *F0*, and *V duration* as fixed factors. VOT is not included in this analysis due to the very limited variations. Significant effects of *F0* [F(2,32) =24.56, p<.01], and *V duration* [F(2,32) =16.55, p<.01] are reported on the proportion of the tense stop, with no significant effect from *Original type* [F(2,32)=3.61, p=.093]. In the aspirated token identification, *Original type* [F(2,32) =11.51, p<.01], F0 [F(2,32) =5.99, p<.05], and *V duration* [F(2,32) =9.49, p<.05] are all found significant. It is interesting that the effect of *V duration* is not very significant in the proportion of the plain stop identification. ANOVA shows a significant effect of *Original type* [F(2,32) =16.79, p<.01] and a highly significant effect of *F0* [F(2,32) =103.93, p<.001], but no robust effect of *V duration* [F(2,32) =3.94, p=.081] in identifying the plain stop. The ANOVA results suggest that the acoustic cues do not role in a uniform way to signal individual stop laryngeal contrasts. It is still confirmed that vowel duration roles as an acoustic cue in identifying the stop contrast, at least the aspirated and the tense in Korean.

Figure 1 further shows interaction of F0 and vowel duration in identifying the stop laryngeal contrast. Regardless of the source type, the answers are found to depend on the edited F0 and vowel duration values consistently in Figure 1, where lines and figures depict answer rates of each stop token with F0 manipulations. For example, the dashed line with a filled circle for ph_160Hz displays the percentage of the aspirated stop identification for the tokens of 160Hz. In Figure1, the tokens with the lowest F0, marked as solid lines, have very high proportion of the plain stop identification, though the rates slightly decrease as vowel duration increases on the x-axis. The overall answer rates for the tense and the aspirated are very low in the lowest frequencies, though the answer rates for the tense slightly increase as the duration become greater. The patterned lines for tokens of the mid and high frequencies, on the other hand, display relatively greater answer rates for the aspirated and the tense, while the answers of the perceived plain stop are very few. The perceptual dominance of F0 in identifying the lax stop [4] is confirmed in the current study, and is enlarged to the dominance over the vowel duration in the current findings. The effect of F0 is further obvious, when the tokens are provided without editing the duration as in Figure 2. Under fixed unambiguous vowel durations, the tense stop shows most frequent identification in the mid F0 ranges, the aspirated in the highest F0, whereas the plain in the lowest F0 ranges.

Figure 1: The percentage of answers in terms of vowel duration and F0.



The tokens with higher frequencies in Figure 1, however, show the separation of the aspirated and tense stops, the two contrastive tokens with the opposite durational patterns, which indicates a significant effect of V duration. Both the tense and the aspirated are marked with higher F0 values, but are found in the opposite ways as the duration changes. The difference is more obvious in Figure 3, where F0 conforms to the corresponding

contrast whereas V durations are edited into three ways. The different V durations correspond to different stop identification under the unambiguous F0. The short V duration leads to a very low identification rate of the tense, whereas the answer rates for the tense become higher when the duration increases. The aspirated shows the opposite patterns as the shortest with the highest answer rates. Figure 3 clearly shows the influence of V duration in identification of stop laryngeal contrast in Korean, taking long duration for the tense and short duration for the aspirated with unambiguous F0s.

Figure 2: The percentage of answers in terms of F0 when the duration is set for the given contrast.

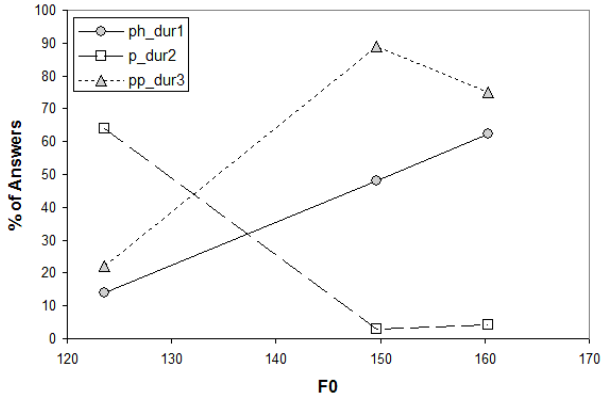
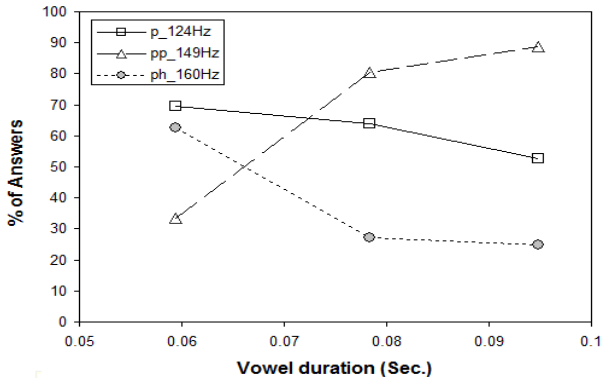


Figure 3: The percentage of answers in terms of vowel duration when F0 is set for the given contrast.



The overall confidence of answer rates is examined with the self-reported confidence level points. Pearson correlation coefficient reveals that the percentage of the most popular answers is significantly correlated with the reported confidence levels ($r=0.606$, $p<.001$) and standard deviation of the confidence levels ($r=-0.587$, $p<.001$). Most common answers are marked with very high confidence from the listeners with less variation in their confidence, confirming the reliability of given answers.

4. DISCUSSION

The results show significant effects of F0 and vowel duration in identifying the preceding stop laryngeal contrast. It is surely confirmed that vowel duration assists the identification of the laryngeal contrast of prevocalic stops in Korean, particularly in the aspirated and plain stops. The temporal resolution, syllable-timing, in Korean, leads to rather uniform syllable durations across various contexts, and a compensatory interaction between VOT and following V durations. The consistency in the vowel duration is suggested as an additional cue for the prevocalic stop laryngeal contrast in the current study. The results also predict plausible language-dependent usage of the acoustic cue. Given that languages use different temporal resolutions, patterns of consonant-to-vowel interaction in a syllable will be dissimilar in languages of different speech timing. In languages with less limited temporal resolutions of syllables, the vowel may not role as a reliable acoustic cue for the preceding stop laryngeal contrast. Unrestricted syllable durations in such languages will not condition compensation between tauto-syllabic consonants and vowels, resulting in less systematic vowel duration in terms of preceding consonantal types. It will be the subsequent study to compare the current findings with the patterns from languages of various speech timing.

5. REFERENCES

[1] Choi, H. Submitted. The durational interaction of vowels and stops under prosodic variation: A cross-language study on English and Korean.

[2] House, A.S., Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Am.* 25, 105.

[3] Kim, C. 1965. On the anatomy of the tensity feature in stop classification: with special reference to Korean stops. *Word* 21, 339-359

[4] Kim, M.-R. Bedder, P.S., Horrocks, J. 2002. The contribution of consonantal and vocalic information to the perception of Korean initial stops. *J. Phonetics* 30, 77-100.

[5] Lisker, L., Abramson, A. 1964. A cross-linguistic study of voicing in initial stops: acoustic measurements. *Word* 20, 384-422.

[6] Umeda, N. 1975. Vowel duration in American English. *J. Acoust. Soc. Am.* 58, 434-445.

[7] Whalen, D., Abramson, A., Lisker, L., Mody, M. 1993. F0 gives voicing information even with unambiguous voice onset time. *J. Acoust. Soc. Am.* 93, 2152-2159.