Correlation between VOT and F0 in the Perception of Korean Stops and Affricates

Midam Kim

Department of Linguistics
Seoul National University, Korea
midam@dreamwiz.com

Abstract
This research examines the trading correlation between VOT and F0 in the production and perception of the three-way distinction of Korean stops and affricates, namely, lenis, aspirated, and fortis, in the word-initial position. For this research, I conducted production and perception tests. For the production test, two female and two male native speakers of Seoul Korean recorded a monosyllabic word list including /ka, kʰa, pa, pʰa, ta, tʰa, ca, cʰa/ 15 times in random order. On VOT-F0 plains, the results showed that lenis, aspirated and fortis were discriminated with the two cues of VOT and F0, without overlapping. The results of a MANOVA test showed that there is a significant difference in lenis, aspirated, and fortis with correlation between VOT and F0 (p<0.001). In the perception test, the stimuli were made by manipulating the sounds recorded in the production test in such a way that F0 values were heightened or lowered at 10 Hz intervals, fixing VOT values. 14 subjects (seven females and seven males) participated in the perception test. The results showed that more than 94% of all fortis stimuli were not influenced by F0 changes, and that VOT and F0 values at the lenis-aspirated boundary showed strong negative correlation (r=-0.923). From these results, I concluded that: 1) Lenis, aspirated, and fortis of Korean word-initial consonants are distinguished with the correlation of VOT and F0; 2) F0 does not function as an acoustic cue in the perception of Korean fortis; and 3) There is a phonetic trade-off between VOT and F0 in the distinction of Korean lenis and aspirated stops and affricates.

1. Introduction

Previous studies on the three-way distinction of Korean obstruents have introduced many acoustic cues, like VOT, the durations of stop closure, following and preceding vowels, intensities of the release burst, and F0 of the following vowels. Among these cues, VOT [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13] and F0 [3], [4], [5], [9], [10], [11], [12], [13], [14], [15], [16] have been considered to be the major factors in distinguishing the three phonation types of Korean stops and affricates in the word-initial position.

In these studies, it has been shown that, in VOT, fortis is the shortest, lenis is middle, and aspirated is the longest, and that, in F0, lenis is the lowest, fortis is middle, and aspirated are the highest. However, in all studies, researchers have reported overlapping of ranges of the three phonation types. VOT often overlapped between lenis and aspirated, and F0 between fortis and aspirated. That is, there existed lenis sounds that had longer VOT values than aspirated sounds, and aspirated sounds that had lower F0 values than fortis sounds. From this, we can conclude that neither VOT nor F0 can be the only independent cue in the three-way distinction of Korean obstruents.

Therefore, we can think of some relationship between VOT and F0. Choi studied the interaction between these two cues [17]. Comparing Chonnam and Seoul Koreans, she examined dialectal variations of the distribution of VOT and F0 in the production of Korean stops. According to the results of the paper, VOT and F0 show different complementary distributions in the two dialects. That is, Chonnam Korean depends more on VOT than on F0, because its salient phrase-initial pitch contour weakens the role of F0 in the stop contrasts, and its phonological contrast in vowel length gives Chonnam speakers more sensitivity to durations of sounds, while Seoul Korean, without such dialectal characteristics, depends more on F0 than on VOT. She showed the complementary interaction between VOT and F0 from dialectal comparisons. However, her main topic was not the interaction itself but the dialectal differences and her finding is not elaborate enough to tell how and how much VOT and F0 interact in the stop contrasts. In addition, her study dealt with only production data. Therefore, we need to study further on the interaction of VOT and F0, not only in production, but also in perception.

Various studies on English [18], [19], [20], [21], [22], [23], [24], [25], [26] showed that acoustic cues related with one phonetic category are not perceived in isolation, rather, “They are perceived together and integrated into a unitary phonetic percept, and there is a phonetic trading relation between two cues of the related cues.” [17] We will confirm this finding in the present study on VOT and F0 of the three-way distinction of Korean stops and affricates, by asking whether VOT and F0 cooperate in distinguishing the three categories and there is a trade-off relation between them.

While the majority of studies about of Korean stops and affricates have focused on production, there have been a few studies focusing on perception. Kim conducted a perception test [28] synthesizing VOT, F0, and amplitude of aspiration noise, but the number of F0 levels (i.e. 100, 125, 150 Hz) used in the study were not enough to detect the characteristics of the boundaries among the three categories. Cho's perception test [29], which examined the vowel portion after VOT, provided the significant finding that only the vowel portion can distinguish the preceding consonant by 67%. However, he did not manipulate individual vowel features, so he could not identify which factor made the vowel portion so important. Han conducted a perception test [8] synthesizing F0 values of the vowel portion, but she used fortis and lenis sounds as her stimuli, excluding the aspirated, giving only a two-way distinction. Lee & Jung carried out a variety of perception tests regarding VOT, closure duration, and the following vowels [11], but they did not conduct a test of F0, because of technical constraints. The perception test in [12] cross-splited VOT and the vowel portion, finding out that the vowel portion carries as much importance as the consonant portion. In addition, she concluded that low F0 values uniquely specify lenis stops. However, her study was constrained by the fact that she did not control for each feature of the vowel portion. So she was unable to identify which feature of the vowel portion is the most important. That is, her conclusion is from the statistical analysis on the production test results, not from her perception test results. Therefore, we need to conduct a perception test of the three-way distinction, in which related features are controlled and the independent variables have sufficient levels.

In recent studies, the voiced portion, especially its F0 value, has attracted more attention [11], [12], [28], [29]. However, manipulation of F0 has not been attempted enough to determine whether the results of the perception tests support the hypotheses that F0 is more important than VOT, nor to detect which factor makes the voiced portion more important than the non-voiced portion [11], [12], [29]. A few perception tests have attempted to manipulate or synthesize F0 [8], [28] and obtained significant results. However, these studies were limited, as stated above, by the fact that Kim did not test a broad range of F0 stimuli enough to show the boundary characteristics [28], and Han dealt with fortis and lenis, that is, only the two-way distinction [8].

The current study focuses on the correlation of VOT and F0 in the
production and perception of the three-way distinction of Korean stops and affricates. This study is organized as follows. The procedure and results of the production test are discussed first. In this test, I measured and analyzed VOT and F0 of the three phonation types of Korean stops and affricates in their word-initial position, focusing on the correlation of the two cues. Second, I discussed the perception test. In the perception test, I manipulated F0 values of following vowels of the sounds recorded in the production test, and conducted an identification test. With the result of the identification test, I examined the characteristics of the categories of the three phonation types and the boundary lines among them. The scope of this study is limited to Seoul Korean.

2. Production test

2.1. Methods

2.1.1. Subjects and Materials

Four native speakers of Seoul Korean (2 females and 2 males, including the female author) participated in the production test. The reading material consisted of 15 repetitions of 12 Korean nonsense words in CV sequences, including /ka/, /kʰa/, /kʰa/, /pa/, /pha/, /pʰa/, /ta/, /tha/, /tʰa/, /ca/, /cha/, /cʰa/. Words in the material were served and read, in isolation, in a random order.

2.1.2. Recording

Recording was made, in a quiet room, digitally with an AKG C1000S microphone through CoolEdit 2000 into a PC equipped with SB Platinum Ex sound card. Sampling rate was set at 22050 Hz. The distance between the microphone and the subject was 5 cm.

2.1.3. Measurements

Each target word was analyzed with Praat 4.0.53. VOT and F0 of the following vowels were measured. In the Spectrogram Settings, Gaussian window was used with 5 ms of window length, 40 dB of dynamic range, and 6 dB/oct of pre-emphasis. In the Pitch Settings, default autocorrelation method of Praat was used with Gaussian window (‘very accurate’ in the menu). The onset of the voiced part of the vowel was located at the nearest zero crossing point of the onset of the first formant of the vowel in the spectrogram. VOT was measured from the release burst to the onset of the voiced part. F0 was taken at the onset of the voiced part.

2.2. Results

In the results, VOTs are the longest in the aspirated, intermediate in the lenis, and the shortest in the fortis. In addition, F0s are the highest in the aspirated, intermediate in the fortis, and the lowest in the lenis. The orders of VOT and F0 values are consistent throughout all four places of articulation, and for all four subjects, but overlapping often occurs. VOT ranges overlap between the lenis and aspirated for all places of articulation for all subjects. F0 ranges also overlap between the aspirated and fortis, among all three phonation groups. Detailed information about overlapped phonation types of VOT and F0 is given in Table 1 and 2.

Table 1: Overlapped phonation types of VOT (L: lenis, A: aspirated, F: fortis) (Subject A, B, C, D).

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Vowel Stop</th>
<th>Bilabial Stop</th>
<th>Alveolar Stop</th>
<th>Palatal Affricate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>L-D</td>
<td>L-A</td>
<td>L-A</td>
<td>L-A</td>
</tr>
<tr>
<td>B</td>
<td>L-A</td>
<td>•</td>
<td>L-A</td>
<td>L-A</td>
</tr>
<tr>
<td>C</td>
<td>L-A</td>
<td>L-A, F-L</td>
<td>L-A, F-L</td>
<td>L-A, F-L</td>
</tr>
<tr>
<td>D</td>
<td>L-A</td>
<td>L-A, F-L</td>
<td>L-A</td>
<td>L-A</td>
</tr>
</tbody>
</table>

The scatter plots of VOT and F0 of all four subjects are illustrated in Figure 1. Note that the overlapping among the three phonation types does not occur in the plot of VOT and F0. Therefore, VOT and F0 seem to contribute to the three-way distinction of Korean stops and affricates, when used together.

![Figure 1: Ranges of VOT and F0](image-url)

To test how much VOT and F0 contribute to the lenis-aspirated distinction together, that is, the significances of the correlation of VOT and F0 in the three-way distinction, a MANOVA test was conducted. The independent variables were phonation type and place of articulation, and the dependant variables were VOT and F0. The significance level was set at 0.05. The results are shown in Table 3.

Table 3: Results of the MANOVA test

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Place of Articulation</th>
<th>Phonation type</th>
<th>Interaction Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>F(6,334) p</td>
<td>Value</td>
<td>F(12,334) p</td>
</tr>
<tr>
<td>A</td>
<td>0.302</td>
<td>43.605&lt;0.001</td>
<td>0.058</td>
</tr>
<tr>
<td>B</td>
<td>0.342</td>
<td>39.580&lt;0.001</td>
<td>0.020</td>
</tr>
<tr>
<td>C</td>
<td>0.453</td>
<td>27.083&lt;0.001</td>
<td>0.050</td>
</tr>
<tr>
<td>D</td>
<td>0.443</td>
<td>27.964&lt;0.001</td>
<td>0.038</td>
</tr>
</tbody>
</table>

In the MANOVA test, in both Place of Articulation and Phonation Type, lenis, aspirated, and fortis showed significant differences with the correlation of VOT and F0, for all four subjects. Therefore, it proves that VOT and F0 contribute to the three-way distinction of Korean stops and affricates when used together, with the statistically meaningful correlation between them.
3. Perception test

3.1. Methods

3.1.1. Stimuli

The materials of the female subject A (the author), recorded for the production test, were manipulated for the perception test. F0 values were manipulated with Praat. First, F0 values of the lenis, the lowest in the production test, were heightened up to the highest F0 value of the aspirated at 10 Hz intervals. Second, F0 values of the aspirated, the highest in the production test, were lowered to the lowest F0 value of the lenis by 10 Hz. Last, F0 values of the fortis, intermediate in the production test, were heightened to the highest F0 value of the aspirated and lowered to the lowest F0 value of the lenis by 10 Hz. The range of F0 manipulation was set in each place of articulation, so that the F0 ranges came to be different in all the places of articulation.

In the process of manipulation, I kept every stimulus have just one F0 value, that is, a monitone in order to reduce unintended variable, F0 contour. Kim [31] showed that Korean domain initial consonants have their own F0 contours following constriction release, for example, aspirated and tense stops/affricates form an H tonal pattern, whereas lax stops/affricates form an LH tonal pattern. We can conclude from this study that F0 contours might be another acoustic cue in the three-way distinction of Korean domain initial obstruents, which are not intended in this study.

VOT was not manipulated because of the technical constraint. VOT is not a homogeneous period, whose spectral character changes constantly. Therefore, we cannot cut off or add certain portion to control its duration. Nor can we compress or expand the duration, which is the method of manipulating duration in Praat. Instead of manipulating VOT, I used every file recorded in the production test, whose range of VOT is large.

Table 4: The numbers and F0 Ranges of Stimuli

<table>
<thead>
<tr>
<th>Original data/Number of stimuli</th>
<th>F0 range (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>kp</td>
<td>222 to 331</td>
</tr>
<tr>
<td>kpʰ</td>
<td></td>
</tr>
<tr>
<td>kpʰ</td>
<td>219 to 358</td>
</tr>
<tr>
<td>kpʰ</td>
<td>225 to 325</td>
</tr>
<tr>
<td>kpʰ</td>
<td>222 to 343</td>
</tr>
<tr>
<td>sum</td>
<td>567</td>
</tr>
<tr>
<td>sum</td>
<td>506</td>
</tr>
<tr>
<td>Total</td>
<td>2014</td>
</tr>
</tbody>
</table>

The total 2014 stimuli were made, which had diverse VOT and F0 values. The list of the numbers and F0 ranges of stimuli is given in Table 4.

Each place of articulation constituted each set, so four sets of stimuli were prepared for the perception test.

3.1.2. Subjects

Fourteen subjects (seven females and seven males) participated in the perception test. All subjects were in their 20s and native speakers of Seoul Korean. No subjects reported any history of hearing disorders.

3.1.3. Procedure

Inquisit 1.33 was used for the tests. A subject was instructed to identify a stimulus by clicking one of the three buttons on the monitor, where Korean orthographies were presented. The response mode was free. In the study four test sets were prepared according to the four sets of stimuli. Each test set was composed of 10 blocks. It took about two minutes for one block, so it took about 20 to 30 minutes to complete a test set. Each test set was done on a separate day, therefore, it took four days to complete the whole tests. Tests were conducted with a Semheiser eH 1430 headphone in a soundproof chamber at the Department of Linguistics, Seoul National University.

3.2. Results

To 717 of all the 816 stimuli, which were manipulated from fortis, all the subjects reported ‘fortis’, that is, 100% of fortis perception. To 99 stimuli, there was disagreement, of which the rates were less than 29%. This indicates that the fortis is not likely to be influenced by changes in F0. Therefore, the remainder of this paper will deal with the distinction between the lenis and aspirated.

Figure 2 shows a typical categorical perception. In these pictures, we can notice that F0 values at the boundary become smaller as VOT values become larger, as shown in Table 5.

Table 5: VOT and F0 values at the boundaries of /ka, kʰa/ distinction. (CB means the categorical boundary.)

<table>
<thead>
<tr>
<th>VOT (ms)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>26.5</td>
<td>302</td>
</tr>
<tr>
<td>30.9</td>
<td>294</td>
</tr>
<tr>
<td>41.5</td>
<td>273</td>
</tr>
<tr>
<td>44.1</td>
<td>269</td>
</tr>
<tr>
<td>69.7</td>
<td>254</td>
</tr>
<tr>
<td>90.4</td>
<td>246</td>
</tr>
</tbody>
</table>

The correlation coefficients of VOT and F0 values obtained at the lenis-aspirated boundaries are given in Table 6.

Table 6: Correlation Coefficients of VOT and F0 at the lenis-aspirated boundaries

<table>
<thead>
<tr>
<th>Obstruents</th>
<th>Correlation Coefficients of VOT and F0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar Stop</td>
<td>-0.923</td>
</tr>
<tr>
<td>Bilabial Stop</td>
<td>-0.957</td>
</tr>
<tr>
<td>Alveolar Stop</td>
<td>-0.95</td>
</tr>
<tr>
<td>Palatal Affricate</td>
<td>-0.942</td>
</tr>
</tbody>
</table>

As shown in Table 6, there is a strong negative correlation between VOT and F0 at the lenis-aspirated boundary for all four places of articulation. The boundary values and their trend lines are presented in Figure 3.
On the trend lines, F0 values are inversely proportional to VOT values in all for places of articulation. From this result, we can find the cause of an overlapping of VOT and F0 ranges between lenis and aspirated, which have been reported many times. Although aspirated has been characterized by ‘high F0, long VOT’, and lenis, by ‘low F0, short VOT’, the result above clearly shows that, in the perception, it is not always true. In the perception, lenis can be higher in F0 than aspirated, while aspirated may be short in VOT than lenis. The cause is that lenis and aspirated have their own categories on the VOT-F0 plain, with the boundary showing a strong negative correlation between VOT and F0.

4. Discussions

Let me explain the results with a diagram. Figure 4 is a diagram of Figure 3. The previous studies have dealt with only A and B, which have ‘long VOT, high F0’ or ‘short VOT, low F0.’ In this perception test, however, new areas have arisen, such as C1, C2, C3, and C4. C1 and C4 are lenis domains, while C2 and C3, fortis domains. Here, new VOT-F0 relations can occur between lenis and aspirated. When a lenis sound in C1 is compared to an aspirated sound in C3, the lenis sound has a higher F0 value than the aspirated sound. If we compare a lenis sound in C4 with an aspirated sound in C2, the VOT value of the lenis sound is longer than that of the aspirated sound. Therefore, it is not that lenis always has lower F0 values or shorter VOT than aspirated.

We can conclude, therefore, that VOT and F0 are in a phonetic trading relation. In the perception, as VOT is getting shorter, F0 ranges of lenis become larger, while those of aspirated come smaller. Inversely, as VOT is getting longer, F0 ranges of aspirated become larger, while those of lenis come to be smaller. This means that VOT and F0 trade off each other and cooperate as a unitary acoustic cue in the three-way distinction of Korean stops and affricates in the word-initial position.

We can put together the results of the production test and the perception test with Figure 5.

Here, we can notice that the results of the production test and the perception test are consistent with each other. The lenis-aspirated boundary lines gained in the perception test lie between lenis and aspirated sounds of the production test. This tells us that the production and perception of the three-way distinction of Korean stops and affricates correspond to each other.

The production test shows that the correlation of VOT and F0 contribute to the three-way distinction of Korean stops and affricates in the domain initial position. The VOT and F0 ranges of lenis, aspirated, and fortis overlap, respectively, but when VOT and F0 are adopted together, the ranges of the three categories do not overlap; therefore, the conclusion is reached that the correlation of VOT and F0 can make a three-way distinction of Korean stops and affricates.

The fortis has been known to have relatively high F0 values, which is also confirmed in this study, but its perception proves to be rarely influenced by F0 values. The reason for that seems to be that the fortis has extremely short VOT, which is a quite different factor from lenis and aspirated. Thus, we can conclude that short VOT is sufficient acoustic cue for fortis.

The negative correlation is clearly found in the categorical boundary lines of the lenis and aspirated in the perception test. It is known that the aspirated has the highest F0 and the longest VOT, and the lenis has the lowest F0 and short VOT. However, she did not describe the detail of how the cues work together. This study showed clearly the way VOT and F0 trade off each other in the contrasts. This is significant in that it may shed light on how cues work together, and it can be introduced in other studies as another acoustic cue.

Phonetic trading relations occur because the cues are phonetically equivalent with respect to the contrast in question [31]. VOT and F0 can be phonetically equivalent with respect to the three-way distinction of Korean stops and affricates. This assumption is opposed to the Kim’s conclusion that F0 of the vowel portion is more important than VOT in the three-way distinction of Korean stops and affricates [28]. In the present study, it is concluded that VOT and F0 are both necessary for the distinction with similar extents. It needs to be studied farther by the manipulation test of VOT, which remains to be studied.

Other cues other than VOT and F0 may have their own relations. I suggest that when we study acoustic cues, we should also study their relations rather than study the cues separately. Through the investigation on cue relations, it will be found new aspects of Korean
three-way distinction of stops and affricates. This study is limited in Seoul Korean, but the relation of VOT and F0 may be different in other dialects of Korean. Especially Gyeongsang Korean seems to be very different, because it is a tonal language, which has different characteristics, compared to Seoul Korean. Of course, Gyeongsang Korean also uses the three-way distinction of stops and affricates. Therefore, it becomes very important to examine the characteristics of VOT-F0 relation in Gyeongsang Korean in later studies.

This study can be used for Korean language education. Nowadays, there are many non-native speakers in Korean whose languages use voiced/unvoiced distinction, that is, the two-way distinction of stops and affricates. For example, English and Japanese have difficulties in making the three-way distinction of Korean stops and affricates. They could improve their Korean speaking and listening performance by reminding the F0 difference between lenis and aspirated. These assumptions should be examined in later studies.

5. Conclusions

The three-way distinction of Korean stops and affricates in the word-initial position can be made by the correlation of VOT and F0 in both production and perception. Fortis has the shortest VOT and a relatively higher F0, but in the perception, it is not likely to be influenced by F0 changes. Lenis has relatively shorter VOT and lower F0, while aspirated has longer VOT and higher F0, with their boundary lines have negative correlations between VOT and F0. In conclusion, there is a phonetic trading relation between VOT and F0 with respect to lenis-aspirated distinction of Korean stops and affricates in the word-initial position. Figure 6 summarizes this.

6. References

[8] Han, Jeong-Im, The Phonetics and Phonology of “Tense” and “Plain” Consonants in Korean, Ph. D dissertation. Cornell University, 1996.
[28] Kim, Mi-Ran, Charles Read, Keith Klunder, & Andrew Lotto, “Production and perception of word-initial stops by Korean adults”, Journal of the Acoustical Society of America, 94, 1865,
1993.

