VOT in the Surface Distinction of Korean Plain and Tense Stops in Initial Position: A Perception Test

Jeong-Im Han*

ABSTRACT

This paper tests whether the VOT (Voice Onset Time) contributes to the perception of tense and plain stops in Korean in the initial position. Previous studies have yielded opposite results regarding the role of the VOT in the distinction of these two stop categories. One the one hand, Lisker and Abramson (1964), Kim (1965), and Han and Weitzman (1970) suggest that the Korean plain and tense stops exhibit overlapping values of VOT and thus that, unlike many other languages, the VOT fails to serve as a cue to separate these stop categories. On the other hand, Silva (1991, 1992), and Lee (1991, 1994) use similar methods, but find that the VOT serve as the cue. In this study, the differences in the acoustic measurement with respect to the VOT of Korean plain and tense stops in initial prevocalic position will be tested preceptually, in seeking to determine whether these acoustic differences are used by listeners to distinguish these two stop types.

Keywords: Voice Onset Time, Korean stops, perception

I. Introduction

Lisker and Abramson (1964) show how languages that have been traditionally described in terms of a phonemic voicing distinction may phonetically realize this distinction differently. In their study of initial stops in eleven languages, Lisker and Abramson (1964) found that languages show systematic differences in the amount of vocal cord vibration and post-release aspiration associated with each stop. They explain these results with language-specific differences in Voice Onset Time (VOT), the time interval between the release and the onset of voicing. This timing difference distinguishes phonetically voiced stops (voicing begins before the stop release) from voiceless unaspirated stops (just after the release) and voiceless aspirated stops (voice onset lags considerably behind the release).

Focusing on the Korean stops, the VOT values of Korean stops measured by Lisker and Abramson are given in Table 1.[1)]

* Language Research Institute, Seoul National University

1) Korean is widely assumed to have a three-way underlying phonation contrast in stops in initial position such as plain (tal 'moon'), tense (t'al 'daughter') and aspirated (t^hal 'mask').

Table 1. Lisker and Abramson's VOT data (ms) of Korean Initial Position

	/p'/	/p/	/p^h/	/t'/	/t/	/t^h/	/k'/	/k/	/k^h/
average	7	18	91	11	25	94	19	47	126
range	0:15	10:35	65:115	0:25	15:40	75:105	0:35	30:65	85:200
number	15	30	21	16	24	12	16	34	12

(Lisker and Abramson, 1964: 397)

Table 1 shows that Korean seems to be an exception to their argument that the VOT can be used as a means of distinguishing all phonemic categories. First, all of the Korean stops are located in the positive half of the voice onset time continuum. Second, the VOT values can distinguish tense and plain from aspirated stops, but tense and plain types cannot be differentiated in terms of the VOT alone, as they exhibit overlapping VOT ranges. Lisker and Abramson suggest that some other acoustic features are necessary to distinguish tense from plain stops.

Similar findings are presented by Han and Weitzman (1970), whose measurements of the VOT are given in Table 2.

Table 2. Han and Weitzman's VOT Data (ms) of Korean in Initial Position

informant 1 (male)

	/p'/	/p/	/p^h/	/t'/	/t/	/t^h/	/k'/	/k/	/k^h/
average	5	27	129	12	33	133	20	62	148
range	0:15	15:45	80:185	5:25	15:80	85:190	13:35	40:400	9:205

informant 3 (female)

	/p'/	/p/	/p^h/	/t'/	/t/	/t^h/	/k'/	/k/	/k^h/
average	5	17	66	6	21	73	15	27	71
range	0:15	5:33	20:95	3:15	10:40	45:130	8:28	15:50	40:110

(Han and Weitzman, 1970: 115)

In Table 2, we observe that differences in the VOT apparently differentiate aspirated from tense and plain stops, but tense and plain stops share a portion of their range in the values of VOT. Kim (1965) also shows similar findings.

Based on these results of acoustic measurements, Han and Weitzman (1970) and Abramson and Lisker (1971) conducted the perception experiments. First, Han and Weitzman (1970) recorded a specific place of aspirated stops followed by vowels [a], [i], and [u]. Based on the wide-band spectrograms made of each syllable, the portion between the stop release and voice onset was cut out in small steps (20 ms) until this time interval was completely removed. Subsequently portions of voicing were cut in 10 ms from the voice onset, for a total of 30 ms. When the created tokens were presented, as more of VOT portions were cut

out, the judgments by subjects switched from aspirated to plain stops, but subjects responded to the stimuli as plain stops even when the VOT was reduced to 1 ms or less, which is within the range of VOT for tense stops. Thus they conclude that VOT cannot distinguish tense from plain stops. Abramson and Lisker (1971) also investigated the perceptual efficacy of VOT for initial stops. They prepared two continua of synthetic VOT variants for the perception tests. One ranged from a voicing lead of 150 ms before the release of the stop to a voicing lag of 150 ms after the release. The other continuum excluded all voicing lead variants. In the response curve for the restricted range, three out of five subjects responded as expected and showed a partition of tense, plain and aspirated in that order. But the remaining two subjects showed only two phonation type responses. The response to the full continuum showed that the three out of five cases, subjects heard any stimulus with the voicing lead as tense stops, while the rest of the continuum was divided between plain and aspirated stops. Thus this study suggests that even though the VOT contributes to distinguish three phonation types in Korean, there must be other characteristics that work with VOT.

Summarizing the results of the both acoustic and perceptual experiments presented so far, the VOT values apparently differentiate tense and plain stops from aspirated stops, but are not sufficient to distinguish the first two categories.

On the other hand, in recent acoustic studies such as Silva (1991, 1992) and Lee (1991, 1994), it was shown that the VOT clearly distinguish tense from plain stops without any overlapping values. Lee (1991)'s data are shown in the following table.

Table 3. Lee (1991)'s VOT Data (ms) of Korean in Initial Position

	/p'/	/p/	/p^h/	/t'/	/t/	/t^h/	/k'/	/k/	/k^h/
average	9	24	78	11	29	75	15	42	89
range	6:13	15:36	59:118	6:14	18:40	49:123	7:23	24:55	54:124

(Lee, 1991: 54)

The mean VOT values of Korean initial stops measured by Lee (1991) appear to be similar to those measured in previous studies. However, the clear difference between this study and previous studies is that Lee's measurements did not show the overlapping VOT values between the plain and tense consonants in each place of articulation. Thus VOT can be used to distinguish these two stop categories, contrary to the claims by previous analyses. Based on the acoustic measurements as in Table 3, Lee argues that

"VOT and the accompanying aspiration is one of the most salient acoustic properties in distinguishing the three stop categories of Korean. Though it has been reported in Kim (1965) that the VOT range of unaspirated consonants [tense consonants] and slightly aspirated consonants [plain consonants] might show overlap, all differences are statistically significant on the one % level in my data" (Lee, 1994: 3).

Silva (1991, 1992) did not explicitly mention that the VOT is sufficient to distinguish

between tense and plain consonants, but he pointed out the fact that his VOT values are noticeably greater than those gathered in previous works. In the following table, Silva's acoustic measurements are presented with those by previous analyses.

Table 4. Mean VOT Values (ms) for Plain Stops in Initial Position

	p	t	k	mean
Silva (1991)	**60**	**51**	**71**	**61**
Kim (1965)	23	38	45	35
Lisker and Abramson (1964)	18	25	47	30
Han and Weitzman (1970)	23	28	52	35

(based on Silva, 1992: 37)

In Table 4, the mean VOT value of the bilabial plain stop is twice greater in Silva's measurements relative to that of the other studies. This large VOT value appears to be sufficient for the distinction of the tense and plain stops. Lee (1991) indicates, however, that it may not be crucial at all whether the VOT value is longer than, or similar to those obtained in the earlier studies: the measurement in Lee shows that the two stop categories are clearly distinguished as their VOT values do not overlap, even when they are small.

In the present study, I employ a simple perception test to reevaluate the role of the VOT in the distinction of tense and plain stops in initial position. The methodology of the experiment used here is different from that of previous perception experiments such as Han and Weitzman (1970) and Abramson and Lisker (1971).

2. Method

2.1 Stimuli

A test continuum was created from computer-edited versions of natural speech. The methodology for creating these stimuli was similar to that used by Miller and Dexter (1987) and Burton, Baum, and Blumstein (1989).

One token each of *pa* and *p'a* was used to create a single *p'a-pa* series. The VOT value of the original *p'a* was 9 ms, while that of the original *pa* was 53 ms.[2] The test continuum was created by replacing increasingly longer segments of the initial portion of *p'a* with segments of equal duration taken from the initial portion of *pa*. The first stimulus was the

2) For the basis of the perception experiment, I measured the VOT values of a small number of tense and plain stops. The present corpus, although small, is adequate for the modeling of stimuli to be used in the perception experiments. The mean VOT values in my acoustic measurement are as follows.

	tense	plain
mean	8	48
range	7-11	28-69
number of tokens	6	6

original tense consonant /p'/. For the second stimulus, the 12 ms of *p'a*, which is the VOT plus some transitional portion between the end of VOT and the beginning of vowel, was replaced by 12 ms of the initial noise part of *pa*. For the third stimulus, the noise part and the first pitch period of *p'a* was replaced by those of *pa*. For the fourth stimulus, the noise part and the first two pitch periods were replaced, and so on, until, for the seventh stimulus, the first 56.4 ms was replaced, which approximately corresponded to the VOT value of *pa*. The overall duration of each CV token was the same. To determine the appropriate step-size, cursors were placed at zero crossings of the vowel waveforms. The duration of noise (burst plus aspiration) and vowel and step size of continuum are presented in Table 5.

Table 5. Duration (ms) of VOT, Vowel, and Step Size for the Test Continuum Members

stimulus	VOT	vowel	step size
1	9		
2	12	117	
3	20.9	108.1	8.9
4	29.7	99.3	8.8
5	38.6	90.4	8.9
6	47.4	81.6	8.8
7	56.4	72.6	9

Stimuli were then transferred to a PC with Bliss software.

2.2 Subjects

Nine Korean native subjects participated in this experiment (six females and three males). They were all graduate students at Cornell University. Subjects had no reported history of hearing impairment.

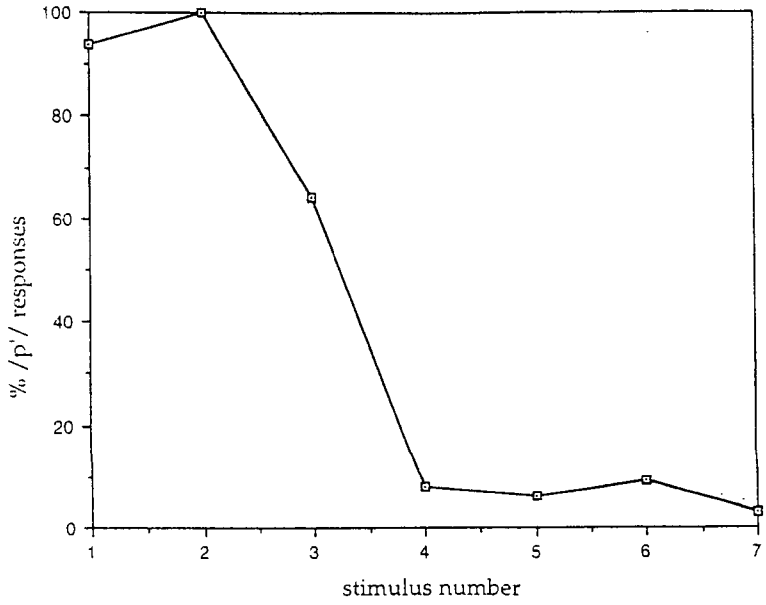
2.3 Procedure

Each subject was tested individually in the Cornell University phonetics laboratory. Subjects were seated in front of a response box containing two response buttons labeled *pa* or *p'a* in Korean orthography. They were instructed to decide whether the sound they heard via headphones (Sony MDR-7506) was *pa* or *p'a*, and to press the appropriate response button. Each continuum was presented 10 times in a random order for a total of 70 stimuli. Only one hand was used for response, and the identification responses were recorded by a PC with BLISS software (Mertus 1989) which controlled this experiment. Before the perception test began, I read prepared instructions to each subject, and ran a set of practice trials. At the end of the test, I debriefed each subject.

3. Results and Discussion

Figure 1 displays the mean identification function for the continuum *p'a-pa,* across subjects. The percentage of /p'/ responses is plotted as a function of VOT. The first stimulus represents the VOT value for the original /p'a/ token. As the figure shows, the /p'/ responses decrease as VOT values increase.

Figure 1. Mean Percentage /p'/ Responses at Different VOT Values. The First Stimulus Represents the VOT Value for the Original Tense Consonant.



The results of the present experiment show that the VOT is an important factor for the perception of tense and plain stops in word-initial position. The curves in Figure 1 show that subjects perceived stimuli categorically, with shorter VOT giving rise to tense percepts and longer VOT giving rise to plain percepts.

This is contrary to the claims of previous analyses such as production studies by Lisker and Abramson (1964), Kim (1965) and Han and Weitzman (1970), and perceptual studies by Han and Weitzman (1970) and Abramson and Lisker (1971), in which the VOT is argued to be insufficient to differentiate tense stops from plain stops in word-initial position. Rather, the results of the perception experiment are in good agreement with the results of acoustic measurements by Lee (1991) and Silva (1991) where these two stop categories are clearly distinguished only with the VOT values without overlapping values between them. Thus this study confirms these acoustic measurements regarding the role of the VOT in the distinction of plain and tense stops, providing perceptual support.

It is of interest that the perceptual range of VOT to be perceived as tense in the results of this experiment was much larger than those in earlier studies such as Lisker and

Abramson (1964) and Han and Weitzman (1970), and even Lee (1991). Figure 1 shows that the first two stimuli whose VOT values were 9 and 12 ms respectively were perceived as tense by almost every subject. The perception of these stimuli does not seem to be a problem in that earlier studies also regard these VOT values as those for tense stops. However, as for the third stimulus, whose VOT values were 20.9 ms, tense consonant responses were around 65 % level, while plain consonant responses were around 35 %. The VOT values of 20.9 ms were considered as those for plain stops in earlier studies, including Lee (1991). For example, in Lee (1991), the VOT range of plain consonants was shown to be 15 through 36 ms, while that of tense consonants was 6 through 13 ms. Thus the stimulus perceived as plain in earlier studies was shown to be perceived as tense in this test. It might be the result of the tendency for the larger value for VOT and its wide range in the results of my acoustic measurements prepared for the perception experiment (and probably Silva's work), compared to earlier works.

Figure 1 appears to show a problem for the analysis I argue in that the /p'/ responses of the first stimulus, which is for the original tense stop token, are a little bit smaller than the synthesized tense stop stimulus (the second stimulus) by around 8%. However, one of the nine subjects is solely responsible for this result: he perceived half of tense tokens as tense, and half of them as plain. All other subjects were shown to perceive this stimulus as tense without any mistake. Thus we can say that it is caused by the single subject's own hearing problem and does not reflect the general hearing (perceptual) pattern of the Koreans.

To make the experiment more complete, two sets of stimuli should be created, not one set as in this experiment. For the present experiment, the test continuum was created by replacing increasingly longer segments of the intial portion of *p'a* with segments of equal duration taken from the initial portion of *pa*. However, another set of stimuli is required to get to the conclusion as above. It should be created by replacing increasingly longer segments of the initial portion of *pa* with segments of equal duration taken from the initial portion of *p'a*. This is, however, not tenable, because it could possibly create stimuli with a portion of aspiration plus a vowel portion plus anther portion of aspiration before the original vowel begins. This could be seen at earlier stimuli, namely the second stimulus through the fifth stiminulus. It can be a problem since it does not sound as natural speeches, especially in Korean stops.

Before concluding this study, let's consider what accounts for the discrepancies in the results of acoustic measurements or perception of VOT between earlier studies and recent ones. There are three possible explanations, the first two of which are proposed by Silva (1992). The first involves the way the VOT is measured. previous analyses and recent ones show a difference in marking the end of VOT in the measurement. Kim (1965), and Lisker and Abramson (1964) measured the period from the beginning of the release to the onset of voicing in the following vowel; and Han and Weitzman (1970) measured from the release to the onset of the first formant. On the other hand, Silva measured from the release to the onset of the second formant, yielding VOT values that are longer, since onset of F2 is

usually later than onset of voicing and F1. Silva suggests that different points marking the end of VOT lead to different VOT values, as shown in Table 5. However, this hypothesis is not tenable when the results of my study are considered, since I measured the VOT from the beginning of the release to the onset of voicing, which si the same measuring criteria as in Lisker and Abramson and Kim. Even though the same criteria as previous studies are used, the VOT values in my measurement are much longer than those in Kim, and Lisker and Abramson, and smaller than Silva's by 12 ms. The VOT values were on average 61 ms followings Silva's way of marking the end of VOT in the measurement, that is, the onset of the second formant. Thus it does not seem to be the case that the differences in measuring criteria can account for the discrepancies in VOT values.

A second possible explanation put forth by Silva is that the value of VOT has increased in Korean native speakers. There is a gap of at least 20 years in the ages of the subjects between previous studies and relatively recent ones such as Silva, Lee and mine, when we consider the age of subjects at the recording. The difference in ages might cause different VOT values, but it is doubtful why the VOT values have changed in recent two years, not before. A more systematic study with a larger sample size is needed to test this hypothesis.

The discrepancies between the present study and previous perceptual studies such as Han and Weitzman (1970) and Abramson and Lisker (1971) might also be due to differences in the methodology.[3] In the present study, the test stimuli were created by replacing the silent interval corresponding to the VOT of original tense and original plain consonants. In contrast, previous studies systematically cut back the aspiration portion of one token of an aspirated consonant. This methodology is less desirable, since it may introduce spectral discontinuities between the manipulated aspiration portion and vowel onset. It is not quite clear at this point if this methodology affects the perception of stimuli, but it seems that the methodology used in the present study contributes to the result of categorical perception of the VOT continuum.

So far, any of the three explanations is not sufficient to account for the discrepancies in the VOT values between earlier work and recent one. A more systematic study is required to fully explain this, which is left for future research.

REFERENCES

이승미. 1991. 한 · 일 파열음의 실험 음성학적 비교연구. *언어학 연구. 12.*

Abramson, A. and L. Lisker. 1971. "Voice Timing in Korean Stops." *Status Report on Speech Research, 27*, July-September.

Burton, M., S. Baum, and S. Blumstein. 1989. "Lexical Effects on the Phonetic Categorization of Speech: The Role of Acoustic Structure." *Journal of Experimental Psychology: Human Perception and Performance, 15*, 567-575.

3) Note that this would only account for the discrepancies in the perception, but not those in the production studies.

Han, M. and R. S. Weitzman. 1970. "Acoustic Features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/." *Phonetica, 22,* 112-128.

Kim, C. W. 1965. "On the Autonomy of the Tensity Feature in Stop Classification." *Word, 21,* 339-359.

Lee, S.-M. 1994. "Aerodynamic and Laryngeal Study of Korean Stops in Word-initial Position." *Paper presented at 127th ASA meeting.*

Lisker, L. and Abramson. 1964. "A Cross-language Study of Voicing Initial Stops: Acoustical Measurements." *Word, 20,* 384-422.

Mertus, J. 1989. *BLISS Manual.* Brown University, Providence, R. I.

Miller, J. and E. Dexter. 1988. "Effects of Speaking Rate and Lexical Status on Phonetic Perception." *Journal of Experimental Psychology: Human Perception and Performance, 14,* 369-378.

Silva, J. 1992. *The Phonetics and Phonology of Stop Lenition in Korean.* Ph.D. dissertation, Cornell University.

Silva J. 1991. "A Prosody-based Investigation into the Phonetics of Korean Stop Voicing." *Harvard Studies in Korean Linguistics, IV,* S. Kuno et al. (editors). Cambridge, MA: Department of Linguistics, Harvard University.

접수일자 : '98. 2. 30.
게재결정 : '98. 4. 10.

▲ Language Research Institute
Seoul National University
Shillim-dong, Kwanak-ku (151-010)
Seoul, 151-742, KOREA
Tel : (02) 880-8573 (O), (02) 534-2577 (H)
e-mail: jihan@plaza.snu.ac.kr