Consonantal and Vocalic Effects in Korean Stop Identification

Mi-Ryoung Kim*

ABSTRACT

This study investigates the contribution of vocalic information following the release of an initial stop to the identification of the three-way stop contrast (aspirated, lax, and tense) in Korean. Recent studies showed that there is a strong interaction between consonant types and tone. The findings raise questions concerning Korean listeners' use of tonal (or vocalic F0) variation in differentiation initial tense, lax, and aspirated stops. The above issues are addressed in the present study using a cross-splicing methodology. The overall results show that low vocalic F0 provided the most salient information for lax stops; tense and aspirated stop identification depended on a combination of VOT, F0, and H1-H2 characteristics. The perceptual dominance of F0 over VOT for lax stops is consistent with the size of the F0 difference in utterance-initial position, as well as their prominent role in Korean intonational phonology.

Keywords: Korean stops, VOT, vocalic F0, tone, cross-splicing, perception

1. Introduction

In utterance-initial position, it is known that Korean stops have a three-way laryngeal contrast for the three places of articulation (e.g., [p*ul] 'horn', [pul] 'fire', [p^hul] 'grass'; [t*al] 'daughter', [tal] 'moon', [t^hal] 'mask'; [k*i] 'meal', [ki] 'energy', [k^hi] 'height'). The stop series are typically described as tense or fortis [p* t* k*], lax or lenis [p t k], and aspirated [p^h t^h k^h] (but see M.-R. Kim, 2000a, b for an alternative analysis of the lax stops). The stops are phonetically voiceless in utterance-initial position. Only the lax type has a voicing counterpart in intervocalic position (e.g., [kibun 'feeling'). A three-way laryngeal distinction among voiceless stops is cross-linguistically atypical. Not surprisingly, then, numerous investigations have explored the precise phonetic characteristics of Korean stops.

VOT (Voice Onset Time; time interval between stop release and the onset of voicing) studies have shown that VOTs are shortest for Korean tense stops, longer for lax, and longest for aspirated (Lisker and Abramson, 1964; C.-W. Kim, 1965; Han and Weitzman, 1970; Hardcastle, 1973; Kagaya, 1974; Silva, 1992; M.-R. Cho Kim, 1994; Y. Kim, 1995; Cho, 1996; Han, 1996a; Shimizu, 1996). At the same time, it is noteworthy that,

* Dept. of English Language and Literature, Korea University

despite systematic mean differences, VOT values for individual tokens of lax and aspirated stops, and lax and tense stops, may overlap (e.g., C.-W. Kim, 1965; Han and Weitzman, 1970; Abramson and Lisker, 1972; M.-R. Cho Kim, 1994; Shimizu, 1996). While differences in VOT values most clearly differentiate Korean aspirated stops from tense and lax stops, differences in F0 best differentiate Korean lax stops from aspirated and tense stops. Although, on average, F0 is lowest for vowels following lax stops, higher following tense stops, and highest following aspirated (Han and Weitzman, 1970; Kagaya, 1974; M.-R. Cho Kim, 1994; Cho, 1996; Han, 1996a; Shimizu, 1996; M.-R. Kim, 2000a), the F0 ranges reported for the tense and aspirated categories substantially overlap.

Korean initial stops also differ in several intensity characteristics, and lax stops emerge as the least intense in terms of release bursts (C.-W. Kim, 1965; Han, 1996a; but see Cho and Keating, 1999, for comparable burst measures of lax and tense stops), aspiration (C.-W. Kim, 1965; Kagaya, 1974; Han, 1996a; but see M.-R. Cho Kim, 1994, for more variable results), and formant structure at the onset of the following vowel (Han and Weitzman, 1970; Hardcastle, 1973). Another vocalic effect is that duration of voicing for the following vowel is longest following tense stops, shorter following lax, and shortest following aspirated (Cho, 1996). Not surprisingly, these relative durations of voiced vowel portions are inversely related to the relative VOT durations of preceding stops. Fiberscopic and cineradiographic studies have also shown that glottal aperture during oral closure is narrowest for Korean tense stops, intermediate for lax, and widest for aspirated (C.-W. Kim, 1970; Hirose, Lee, and Ushijima., 1974; Kagaya, 1974; Jun, Beckman, and Lee, 1998). Kagaya (1974) reported that apertures at voice onset were comparable for the lax and aspirated stops, but different from the complete contact at release/voice onset for tense stops. Recently, Ahn (1999) applied a normalized H1-H2 measure (adjusting for formant frequency effects) to Korean stops. His findings generally consistent with Kagaya's aperture data: the normalized H1-H2 difference at vowel onset was larger for aspirated and lax stops than for tense stops.

Clearly, multiple temporal and spectral properties differentiate Korean tense, lax, and aspirated stops. For initial stops in CV sequences, some of these properties fall within the consonant portion of the syllable while others fall outside conventional stop boundaries, within the (beginning at voicing onset) portion. The current experiments were designed to explore the relative contribution of consonantal and vocalic cues to Korean listeners' perception of word-initial stops.

Previous perceptual studies have shown that variation in VOT alone is not sufficient to elicit a systematic three-way distinction in listeners' judgments, although which phonation type--tense or lax--is poorly identified when VOT is manipulated varies across listeners and stimulus sets. Abramson and Lisker (1972), using a synthetic continuum ranging from 0 to +150 ms VOT, found inconsistent response patterns across five listeners: although three Korean listeners heard a three-way distinction, one listener heard no lax consonants, and another listener unexpectedly reported tense consonants for the middle of the continuum. Poor identification of lax consonants for a synthetic VOT continuum was also reported by M.-R. Cho Kim (1994) (see also Han and Weitzman, 1970). Along similar lines, Han (1996a,b) created a VOT continuum by replacing increasingly longer onset portions of an original /p*a/ stimulus with successively more aspiration from lax /pa/, and found that listeners' identifications shifted from tense to lax when the burst plus first two pitch periods of /p*a/ were replaced. Thus the results of VOT studies suggest that both consonantal and initial vocalic information is important to perception of the three-way phonation contrast. Cho (1996) looked more globally at the contribution of vocalic information, asking Korean listeners to identify initial stops on the basis of syllable fragments in which all aperiodic information plus the first two pitch pulses of the vowel had been excised from original stop-vowel syllables. Across phonation types, identification accuracy was 67%, but accuracy for aspirated stops, which were often misperceived as tense, was just below 50%.

Vocalic information may not be secondary to VOT in the perception of Korean stops, however. On the one hand, as already noted, variation in VOT alone has been shown to be unsuccessful in eliciting the three-way distinction for some Korean listeners; on the other hand, certain effects of stop phonation type on following vowels are particularly strong in Korean. Most notably, the effects of phonation type on F0 in Korean extend well beyond the localized F0 perturbation observed for English and many other languages. M.-R. Kim (2000a), for example, reports systematic F0 effects of word-initial consonants in Korean that spanned several syllables. Acoustic analyses of three-syllable words produced by 8 Korean speakers showed that, although F0 differences between initial aspirated and tense consonants disappeared over the course of the first syllable, low F0 after lax consonants remained: mean F0 was 25 Hz lower after lax consonants than after aspirated or tense consonants in the second syllable, and 15 Hz lower in the third syllable. However, although these F0 effects clearly emerge in word-initial position for isolated words, they should not be viewed as lexical. As shown by Jun (1993, 1996, 1998), the H tone for aspirated and tense stops and the L tone for lax stops occur in when the stops are in phrase-initial, rather than word-initial, position and play a major role in Korean intonational phonology.

In the current study, I tested the importance of vocalic information to the perception of word-initial voiceless stops in Korean. This experiment investigated which cues provide the more robust information about stop phonation type, cues provided by the consonant portions (release burst plus any aspiration) or the vowel portions (beginning at periodic voicing) of the syllable. Listeners were asked to identify the initial consonants of manipulated syllables in which the consonant and vowel portions of naturally produced syllables differing in phonation type had been cross-spliced (e.g., the release burst and aspiration for [p^ha] was spliced onto the vowel portion from [pa]). I expected, for example, that a vowel portion with a low F0 would provide a perceptually salient cue for a lax stop, but the central question was whether such cues would override conflicting information in the consonant portion, such as a strong burst or lengthy aspiration.

2. Methods

2.1 Original Materials

The original stimuli were produced by a 28-year old female native speaker of the Seoul dialect who was a graduate student at the University of Michigan at the time of recording. She read three randomized lists, written in Korean Hangul orthography, at a natural speaking rate. Each list contained 12 monosyllabic target words embedded in the carrier sentence [ige ____ (i)da] 'This is ___' (The initial vowel of the final word in the carrier was pronounced when the preceding target words ended in a consonant). The target words, given in Table 1, were balanced for phonation types (tense, lax, and aspirated), place of articulation (labial, alveolar, and velar), and vowel context (/a/ and /i/), except that the vowel manipulation was restricted to the alveolar context. The place and vowel manipulations were included to test the generality of the perceptual findings.

The recordings were made using an AKG C451 EB condenser microphone and a Panasonic Digital Audio Tape Deck SV-3500 in a sound-attenuated studio in the Phonetics Laboratory at the University of Michigan. The speaker was asked to read the words in a natural speaking voice as if talking to someone at a normal rate. All recorded utterances were digitized at 11.025 kHz and selected acoustic measures were taken using Sound Scope/16 (GW Instruments) speech analysis software. Table 1 gives the three temporal measures, averaged across the three tokens of each stimulus type (ranges are also given). VOT was measured from the release burst to the onset of periodicity in the waveform. Vowel duration was measured from onset of the periodic waveform to onset of the lower-amplitude nasal murmur for nasal-final syllables, and to offset of the periodic waveform for vowel-final syllables. Syllable duration was measured from the stop release to the offset of the syllable. F0 of rime onset, midpoint, and offset was also measured for all stimuli.

Table 1. Mean values for VOT, vowel duration, syllable duration in ms and F0 at onset, midpoint, and offset in Hz. (Duration and F0 ranges for the three averaged tokens are also given)

	Tense		Lax		Aspirated	
Labials	**p*aŋ**	**'bread'**	**paŋ**	**'room'**	**p^{h}aŋ**	**'bang'**
Duration (ms)	mean	range	mean	range	mean	range
VOT	9	6-10	44	35-53	73	62-87
Vowel	168	166-171	157	143-177	110	100-117
Syllable	337	322-355	369	337-388	339	327-356
F_0 (Hz)	mean	range	mean	range	mean	range
Onset	239	225-262	198	197-200	246	242-253
Midpoint	241	232-256	204	202-206	236	232-242
Offset	236	225-248	205	202-208	229	223-232
Velars	**k*aŋ**	**'percussion cap'**	**kaŋ**	**'river'**	**k^{h}aŋ**	**'Mr. k^{h}ang'**
Duration (ms)	mean	range	mean	range	mean	range
VOT	17	13-19	58	47-80	82	63-94
Vowel	136	130-141	131	120-141	118	110-123
Syllable	299	289-314	342	328-340	346	342-355
F_0 (Hz)	mean	range	mean	range	mean	range
Onset	252	252-253	224	218-232	273	271-276
Midpoint	253	248-256	216	210-225	250	248-251
Offset	243	240-243	211	208-216	247	245-248
Alveolars+aŋ	**t*aŋ**	**'ground'**	**taŋ**	**'party'**	**t^{h}aŋ**	**'boom'**
Duration (ms)	mean	range	mean	range	mean	range
VOT	6	4-8	25	19-30	64	63-66
Vowel	155	148-162	137	135-138	94	89-98
Syllable	305	274-338	312	288-325	303	292-308
F_0 (Hz)	mean	range	mean	range	mean	range
Onset	241	237-245	203	200-206	263	259-272
Midpoint	242	240-245	204	199-206	248	245-253
Offset	241	237-245	209	208-212	252	242-259
Alveolars+i	**t*i**	**'belt'**	**ti**	**'D'**	**t^{h}i**	**'dust'**
Duration (ms)	mean	range	mean	range	mean	range
VOT	5	4-6	30	14-34	80	72-86
Vowel	313	293-337	337	325-356	340	307-394
Syllable	318	299-342	359	342-390	359	419-466
F_0 (Hz)	mean	range	mean	range	mean	range
Onset	261	248-272	230	227-232	284	282-286
Midpoint	259	253-266	232	227-235	269	266-272
Offset	254	248-259	231	227-235	257	251-263

VOT values show a clear three-way distinction: longest for aspirated, intermediate for lax, and shortest for tense consonants for most cases. However, similar to previous results reported by C.-W. Kim (1965), Han and Weitzman (1970), and Y. Kim (1995), the stimuli employed in this study do show some overlapping VOT values between lax [k] and aspirated [k^h] (as well as those of [p^h] and t^h]). Although the ranges following labial and velar lax and tense stops substantially overlapped, as did those following velar lax and aspirated, duration of /a/ was shortest after aspirated stops, longer after lax, and longest after tense, consistent with the findings of Cho (1996). For syllable duration, syllables with lax onsets are the longest and those with aspirated ones are the shortest for labials and alveolars followed by /a/. However, syllables with aspirated stops are the longest and those with tense ones are the shortest for velars followed by /a/ and alveolars followed by /i/.

This indicates that syllable durations are not consistent for phonation types across the three places of articulation.

F0 values differentiate the stops, with the effect being strongest at rime onset. At rime onset, F0 values are highest after an aspirated, intermediate after a tense, and lowest after a lax stop, consistent with the findings of previous studies (Han and Weitzman, 1970; Kagaya, 1974; M.-R. Cho Kim, 1994; Y. Kim, 1995; Cho, 1996). However, contrary to Han and Weitzman's (1970) observation, no overlapping of the value ranges for F0 between lax and tense was observed. At midpoint and offset, the F0 values following aspirated and tense are nearly merged, but both are higher than those following lax consonants, consistent with M.-R. Kim's finding (2000a).

Vowel onsets following the tense, as opposed to aspirated and lax, stops also differed in the amplitude of the fundamental or first harmonic relative to higher harmonics. Kagaya (1974), for example, observed that, while glottal aperture at stop release is greater for aspirated than for lax stops, apertures at voice onset for the lax and aspirated stops are comparable, but different from the complete contact at release/voice onset for tense stops. An expected spectral consequence (e.g., Stevens, 1998:425; see also Ladefoged, Maddieson, and Jackson, 1988) is a particularly dominant first harmonic, relative to higher harmonics, for the lax and aspirated stops. This effect is captured for the original stimuli by the difference between the amplitudes of the first and second harmonics (H1-H2), given in Table 2 for each phonation type and stop place.

Table 2. Mean difference (in dB) in the amplitudes of the first and second harmonics (H1-H2) at vowel onset following tense, lax, and aspirated stops. (H1-H2 ranges for the three averaged tokens are given in parentheses.)

	Tense C*	**Lax C**	**Aspirated C^h**
Labial	-6.2 (-6.6 — -6.0)	1.2 (-0.9 — 3.5)	4.3 (2.1 — 6.0)
Velar	-6.8 (-8.2 — -6.0)	4.0 (3.5 — 4.8)	.2 (-4.3 — 3.3)
Alveolar	-4.1 (-4.9 — -3.5)	6.4 (3.5 — 8.8)	2.1 (1.1 — 3.7)
Alveolar+i	-1.0 (-4.0 — 1.3)	5.5 (1.3 — 8.7)	9.3 (8.7 — 10.4)

Measures were based on the FFT spectrum of a 25 ms window beginning at vowel onset. For vowel onsets following lax and aspirated stops, H1 was greater than H2 (with the exception of one token each of /p*/ and /k^h/). The reverse low-frequency spectral tilt held for vowel onsets following tense stops (with the exception of one token of [t*i]).

2.2 Same- and Cross-spliced Stimuli

Same-spliced stimuli refer to stimuli whose C and V portions came from two different tokens of the same phonation type; cross-spliced stimuli refer to stimuli whose C and V portions came from tokens differing in phonation type. In order to determine which portion, consonantal or vocalic, plays a greater role in the identification of Korean stops, all 36 original CV(C) syllables were segmented into consonantal (= burst + aspiration) and vocalic (= rime of a syllable) portions at the onset of voicing, defined as the first pulse of a vocalic waveform that shows features typical of a vowel. Splice locations were identified using a combination of waveform and wide-band spectrographic displays. Measurement accuracy was estimated to be about 0.5 ms (= one cursor movement).

The segmented C and V portions were then same-spliced and cross-spliced within each place of articulation, creating 72 stimuli from each of the original triplets, as illustrated in Table 3 for labials.

Table 3. Examples of same-spliced and cross-spliced stimuli for labials. (Bold stimuli are same-spliced and all others are cross-spliced)

	V*1	**V*2**	**V*3**	**V1**	**V2**	**V3**	**V^h1**	**V^h2**	**V^h3**
p*1		**p*1V*2**	**p*1V*3**	p*1V1	p*1V2	p*1V3	p*1V^h1	p*1V^h2	p*1V^h3
p*2	**p*2V*1**		**p*2V*3**	p*2V1	p*2V2	p*2V3	p*2V^h1	p*2V^h2	p*2V^h3
p*3	**p*3V*1**	**p*3V*2**		p*3V1	p*3V2	p*3V3	p*3V^h1	p*3V^h2	p*3V^h3
p1	p1V*1	p1V*2	p1V*3		**p1V2**	**p1V3**	p1V^h1	p1V^h2	p1V^h3
p2	p2V*1	p2V*2	p2V*3	**p2V1**		**p2V3**	p2V^h1	p2V^h2	p2V^h3
p3	p3V*1	p3V*2	p3V*3	**p3V1**	**p3V2**		p3V^h1	p3V^h2	p3V^h3
p^h1	p^h1V*1	p^h1V*2	p^h1V*3	p^h1V1	p^h1V2	p^h1V3		**p^h1V^h2**	**p^h1V^h3**
p^h2	p^h2V*1	p^h2V*2	p^h2V*3	p^h2V1	p^h2V2	p^h2V3	**p^h2V^h1**		**p^h2V^h3**
p^h3	p^h3V*1	p^h3V*2	p^h3V*3	p^h3V1	p^h3V2	p^h3V3	**p^h3V^h1**	**p^h3V^h2**	

Of the 72 stimuli, 18 stimuli were same-spliced and 54 stimuli were cross-spliced. For instance, the cross-spliced stimuli had the consonantal portion of one token spliced onto the vocalic portion of a token differing in phonation type, but with the same place of articulation. The only constraint on splicing C and V portions was that original syllables (e.g., /p^h/ and /V^h/ from the same token) were not recreated. There were four test sequences (labials, velars, alveolars before /a/, and alveolar before /i/). Each test sequence consisted of 5 randomizations of the 72 stimuli. As a result, a total of 1,440 same- and cross-spliced stimuli (72 stimuli × 5 randomizations × 4 sets) were created.

2.3 Listeners and Procedures

Listeners were 12 native speakers (five males and seven females) of the standard Seoul dialect of Korean. They ranged from 26 to 32 years of age with a mean age of 28.3 years. At the time of testing they were living in Ann Arbor, Michigan; the length of residence in the US ranged from 3 months to 4 years (mean of 2.2 years). In Ann Arbor, which has a large Korean community, they lived in a Korean-speaking environment and reported that they spoke Korean at least 70% of the time. All were paid for their participation. They reported no hearing difficulties and lacked training in phonetics.

For each trial, file sequences were created for the identification tests and output to digital audio tape. Participants were tested in groups of two or three in a sound-attenuated room in the Phonetics Laboratory at the University of Michigan. Participants listened over headphones (Sennheiser HD 265) to stimuli presented via Panasonic Digital Audio Tape Deck SV-3500 at a comfortable loudness level. For each trial, they were asked to identify the syllable they thought they heard and asked to guess if necessary. There were three response options for each trial--the so-called aspirated, lax, and tense--and listeners circled their decision on the answer sheet written in Korean orthography. Prior to the test, the experimenter read the instructions in Korean and presented a familiarization series consisting of 10 same-spliced and cross-spliced practice trials for each set. Each listener responded to 360 same- and cross-spliced stimuli in each of the 4 (labial, velar, alveolar before /a/ and alveolar before /i/) test sequences. In each test, the interstimulus interval was 2 seconds. The end-of-page interval was 10 seconds. A short break was given between each of the 4 sets. The entire testing session lasted about 2 hours.

2.4 Statistical Design

For the analysis of the identification data, all responses were submitted to a Binomial test. The Binomial test is often used to test the validity of the hypothesized probability of an event, when only two events are possible (see William and Beaver, 1994). In an experiment consisting of many trials, with two possible outcomes in each trial, a Binomial test can assess the validity of the hypothesized probability of one of the outcomes, given the sample data. In this case, a Binomial test was used to see which portion, either the consonantal or vocalic portion, played a greater role in differentiating the three stops.

For each same-spliced or cross-spliced stimulus played, listeners could respond either aspirated (= C^hV^h), lax (= CV), or tense (= C*V*). Each stimulus response was classified as (1) both (consonant and vowel) correct, (2) consonant correct, (3) vowel correct, or (4) both incorrect. Out of the four possibilities, the outcomes 'consonant correct' and 'vowel correct' were statistically compared. If a stimulus was same-spliced, only 2 response types--(1) or (4)--were possible outcomes. However, if a stimulus was cross-spliced, three response types were possible: (2) consonant (C) correct, (3) vowel (V) correct, or (4) both incorrect, as shown in Table 4. Note that, because the laryngeal cues carried by the consonant and vowel portions of each cross-spliced stimulus were in conflict (i.e., the consonant and vowel portions contained information about different laryngeal types), no single response type could be encoded as 'correct'. For responses labeled 'C correct', the consonantal portion played a greater role in stimulus identification; similarly, the vocalic portion played a greater role for responses labeled 'V correct'.

Table 4. Response types--'C correct', 'V correct', and 'both incorrect'--for the identification of cross-spliced stimuli

stimulus	C correct	V correct	Both incorrect
C*V	tense	lax	aspirated
$C^{*}V^{h}$	tense	aspirated	lax
CV*	lax	tense	aspirated
CV^{h}	lax	aspirated	tense
$C^{h}V^{*}$	aspirated	tense	lax
$C^{h}V$	aspirated	lax	tense

The responses to the same-spliced stimuli were not submitted to a Binomial test because 9 out of 10 stimuli were correctly identified. Responses to the cross-spliced stimuli were tested for the portion -consonant correct or vowel correct- which played a greater role in listeners' identification of the three phonation types.

3. Results

Prior to detailed discussion of the results, I present a confusion matrix in Table 5 that shows the types of errors(%) listeners made in their responses to each of the same-spliced and cross-spliced stimulus types. Table 5 gives the mean responses of 12 listeners.

Table 5. Pooled percent tense (C*), lax (C), and aspirated (C^{h}) responses of 12 Korean listeners to same-spliced and cross-spliced stimuli (averaging across place of articulation and vowel context) Number in underlined = C portion correct; number in bold = V portion correct.

Stimulus	Type	Response Type		
		C* (tense)	C (lax)	C^{h} (aspirated)
Same-spliced	C*V*	99	0	1
	CV	0	98	2
	$C^{h}V^{h}$	1	1	98
Cross-spliced	C*V	7	**92**	1
	$C^{*}V^{h}$	78	5	**17**
	CV*	**66**	6	30
	CV^{h}	24	6	**70**
	$C^{h}V$	0	**81**	19
	$C^{h}V^{*}$	**23**	0	77

As seen in Table 5, responses to the same-spliced stimuli (tense C*V*, lax CV, and aspirated $C^{h}V^{h}$) were nearly always correct. As can be expected from the 98-99% accuracy, there was essentially no variation across place of articulation, vowel context, or listeners (except that the accuracy of two listeners fell below 80% correct on the same-spliced lax stimuli, [paŋ]). Our auditory assessment of the manipulated stimuli was that splicing itself did not degrade the quality of the stimuli, and this is supported by listeners' responses to these control stimuli.

Accuracy of responses to the cross-spliced stimuli ranged from 17-92% correct for the vocalic portion and 6-78% correct for the consonantal portion; vocalic influences predominated for four cross-spliced stimuli (CV*, C*V, C^hV, and CV^h) and consonant influences for the remaining two ($C*V^h$ and C^hV*). In the next section, these patterns are broken down according to consonantal place and vowel quality.

In analyzing listeners' responses to the cross-spliced stimuli, which consisted of conflicting cues, I consider whether the consonantal or vocalic portion played a greater role in stop identification. I present the overall pooled-results in section 3.1, including place, vowel, and phonation effects. Individual effects are presented in section 3.2.

3.1 Pooled Results

Listeners' responses to cross-spliced stimuli consisted of three types: 'C correct', 'V correct', and 'both incorrect' (see Table 4). Recall that for a cross-spliced stimulus C*V, the label 'C correct' was applied when a tense type (C*V*) was chosen, 'V correct' when a lax type (CV) was chosen, and 'both incorrect' when an aspirated type (C^hV^h) was chosen. As seen earlier in Table 5, for four of the six cross-spliced stimuli, 'V correct' responses predominated. The overall mean percent choice of the three response types to the cross-spliced stimuli averaged across 12 listeners was 58% for the 'V correct', 32 % for the 'C correct', and 10% for 'both incorrect'. Listeners were nearly twice as likely to base their responses on the vocalic information (7,519 responses out of 12,960) as on the consonantal portion (4,119 responses). They were least likely to choose a response that did not match either the vocalic or consonantal portion (1,322 responses). Binomial tests showed that the vocalic effect was significantly greater than the consonantal effect ($p < 0.0001$). The overall results indicate that the vocalic portion played a greater role than the consonantal portion in listeners' identification of the three stop types.

The greater effect of the vocalic than the consonantal portion held across the three places of articulation, as shown in Table 6. Binomial tests showed that the vocalic effect was significantly greater than the consonantal effect for each place of articulation ($p < 0.0001$). Although highly significant for all places, the preference for the vocalic portion was slightly weaker for labials, due largely to the relatively high percentage of 'both incorrect' labial choices.

Table 6. Pooled percent responses categorized as 'C correct', 'V correct', and 'both incorrect' according to place of articulation. The number of responses in each category is given in parentheses. Cases that are significantly greater ($p < 0.05$) are ***in bold***.

Response type	Labial	Alveolar	Velar	Sum
C correct	32% (1025)	27% (870)	29% (952)	29% (2847)
V correct	**54% (1747)**	**66% (2153)**	**65% (2104)**	**62% (6004)**
both incorrect	14% (468)	7% (217)	6% (453)	9% (869)

The greater effect of the vocalic portion on stop identification also held for both vowel contexts, as seen in Table 7. Binomial tests showed the vocalic effect to be greater than the consonantal effect for both the /a/ and /i/ contexts ($p < 0.0001$ in both cases). However, listeners' preference for vocalic information was not nearly as robust for the /i/ context (47%) as for the /a/ context (66%). I return to this point later in this paper.

Table 7. Pooled percent responses categorized as 'C correct', 'V correct', and 'both incorrect' according to vowel context. The number of responses in each category is given in parentheses. Cases that are significantly greater ($p < 0.05$) are ***in bold***.

Response type	Alveolar+$a_ŋ$	Alveolar+i
C correct	27% (870)	39% (1272)
V correct	**66% (2153)**	47% (1515)
both incorrect	7% (217)	14% (453)

Next, consider the greater influence of vocalic information held for all combinations of phonation types. Binomial tests showed that vocalic effect was significantly greater than consonantal effect for all three consonantal phonation, tense (C*), lax (C), and aspirated (C^h). Vocalic effect was also significantly greater than consonantal effect when vocalic phonation was lax (V) and aspirated (V^h) ($p < 0.0001$). However, consonantal effect was significantly greater than vocalic effect when vocalic phonation was tense (V*) ($p < 0.0001$). The statistical results indicate that vocalic effects did not hold for all phonation types. Phonation variation can be seen in Table 8 where, of the six cross-spliced stimulus types, four stimuli (C*V, CV*, CV^h, and C^hV) had a greater vocalic than consonantal influence, and two stimuli ($C{*}V^h$ and $C^hV{*}$) had the opposite pattern.

Table 8. Pooled percent response categorized as 'C correct', 'V correct', and 'both incorrect' according to each cross-spliced stimulus. The number of responses in each category is given in parentheses. Cases that are significantly greater ($p < 0.05$) are ***in bold***. (where C = [p, t, and k] and V = [a] and [i])

Cross-spliced stimuli	Response categorization	Mean	Phonation type : perceived as...
C*V	C correct	7.3 (118)	tense
	V correct	**92 (1493)**	**lax**
	both incorrect	0.6 (4)	aspirated
C^*V^h	**C correct**	**71 (1158)**	**tense**
	V correct	23 (366)	aspirated
	both incorrect	5.9 (96)	lax
CV*	C correct	26 (418)	lax
	V correct	**58 (936)**	**tense**
	both incorrect	16 (266)	aspirated
CV^h	C correct	6.5 (105)	lax
	V correct	**85 (1380)**	**aspirated**
	both incorrect	8.3 (135)	tense
C^hV^*	**C correct**	**69 (1125)**	**aspirated**
	V correct	30 (489)	tense
	both incorrect	0.4 (6)	lax
C^hV	C correct	8.9 (144)	aspirated
	V correct	**83 (1342)**	**lax**
	both incorrect	8.3 (134)	tense

Three patterns can be extracted from the results in Table 8. First, the vocalic effect is robust when the phonation type of the vocalic portion is lax. As shown in Table 8, listeners' responses largely relied on the vocalic portion for the perception of C*V and C^hV. Despite the conflicting cues between the consonantal and vocalic portions, listeners' identified these stimuli as lax without much confusion. Second, when the consonantal portion was lax, but the vocalic portion was not (CV* and CV^h), the vocalic portion overrode the consonantal portion. However, for CV*, the relatively low percentage of vocalic responses (58%) indicated that listeners were confused. Third, the consonantal portion was crucial whenever aspirated and tense types were cross-spliced (C^*V^h and C^hV^*). Given that the C^*V^h and C^hV^* stimuli had a similar H tone, it is reasonable that the amount of aspiration in the consonantal portion played an important role in distinguishing tense from aspirated types (see section 4 for more discussion).

The observations generally held across consonantal places and vocalic contexts, albeit with some variation, as shown in Figure 1. Exceptions are pV* and tV^h (V=/i/). For the identification of pV*, neither the consonantal nor vocalic portion played a large role because listeners mostly heard pV* as having an aspirated consonant. Similarly, for tV^h, listeners mostly reported hearing a tense consonant. Listeners rarely perceived these stimuli as having lax consonants. This finding indicates that tense syllables can be confused with aspirated ones due to the H tone, but lax syllables are not confused with tense or aspirated because of their tonal difference. A similar interpretation can be applied to other stimuli whose consonant is lax. Tone does not explain, however, why the responses to pV* differed from those to tV* and kV*, nor why the responses to tV^h differed from those to pV^h and kV^h. In addition, listeners presumably used non-tonal information for the distinction between H-toned CV* and CV^h.

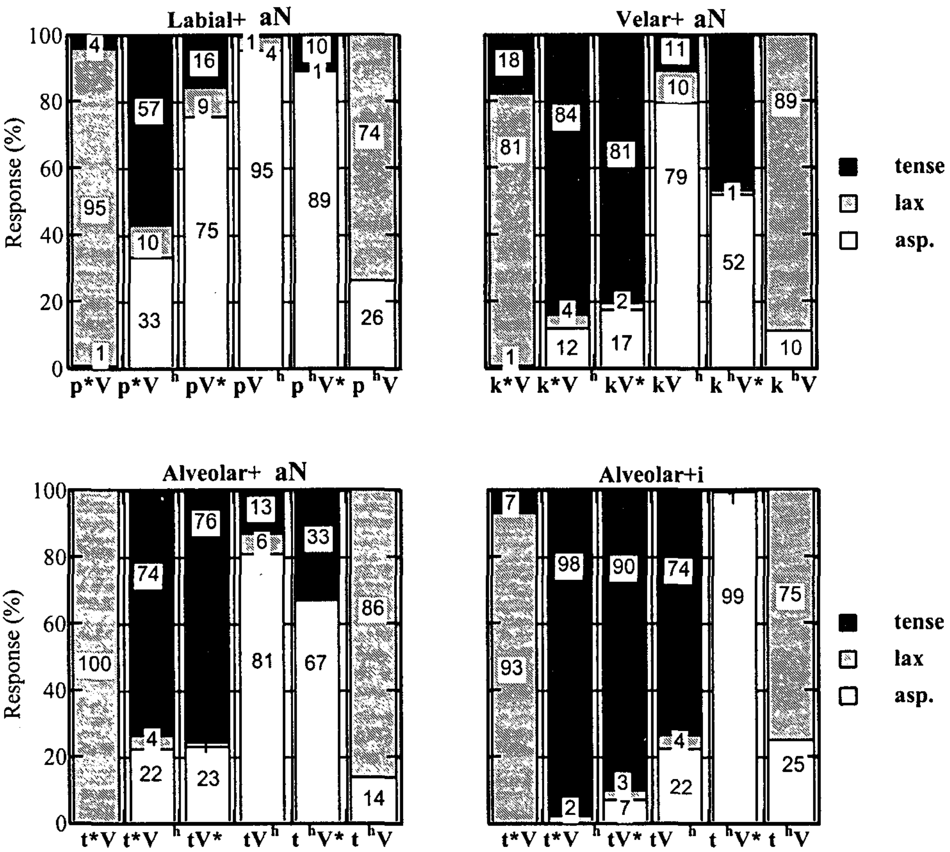

Figure 1. Pooled response rates (%) of the three phonation types according to stimulus type (C*V, $C*V^h$, CV*, CV^h, C^hV*, and C^hV for the three places of articulation followed by /-aŋ/ and alveolar followed by /-i/. (where N = velar nasal)

The acoustic characteristics of the test stimuli given in Table 1 suggest some non-tonal factors that may have influenced listeners responses. For the responses to CV* and CV^h, given that the consonantal portion is lax in both cases, and the tone of vocalic portion is the same (H), it was expected that listeners would be confused in distinguishing these syllables. Surprisingly, for all CV* and CV^h stimulus types except for pV* and tV^h, listeners were able to correctly identify the vocalic portion more than half of the time. Two speculations can be made: One is that listeners may use vowel duration as a secondary cue for the distinction, as pointed out by Cho (1996). Recall that, all else being equal, vowel duration is in general longest for tense syllable types and shortest for aspirated syllable types. This is true for the current data except for a few cases (see Table 1). The other speculation is that listeners may use the amount of aspiration for a specific lax consonant. Note that the VOT for lax consonants in Table 1 varies from one token to the other: 44 ms for [p], 28 ms for [t], and 58 ms for [k] (where vowel =[a]). To some extent, the longer the aspiration, the more accurate the identification of CV^h stimuli. However, since neither factor provides a satisfactory account for all the responses to CV* and CV^h across places and vowels, it seems that the cues of vowel duration and aspiration together may have helped listeners to distinguish between tense and aspirated consonants.

3.2 Individual Results

Table 9 gives the mean percent of each response type for each listener, summing up all stimuli. Eleven listeners were primarily influenced by the vocalic portion of the syllable; only for listener CP was the consonantal influence slightly greater than the vocalic influence. However, there was enormous cross-listener variation, with the overall contribution of the vocalic portion ranging from 42% to 92%. This large variation among listeners indicates that different listeners employed different strategies in speech perception. Listeners SC, JP, CP, JK and KP had considerable difficulty in identifying stop types on the basis of the vocalic portion. In addition, certain stimuli--such as pV* and k^hV*, tV*, t^hV followed by /i/--were highly variable across listeners.

Table 9. Individual percent responses categorized as 'C correct', 'V correct', and 'both incorrect' averaged across the three places of articulation followed by /a/. Cases that are significantly greater ($p < 0.05$) are ***in bold.***

Listeners' responses

Response type	**CP**	**EP**	**JK**	**JC**	**HK**	**KP**	**YR**	**SC**	**YY**	**SM**	**JP**	**YP**
C correct	48	20	34	30	6	38	19	37	30	27	40	30
V correct	42	**77**	**59**	**63**	**92**	**57**	**75**	**55**	**63**	**66**	**51**	**60**
both incorrect	10	3	7	7	2	5	6	8	7	7	9	10

4. Discussion and Conclusion

The results show that, when presented with conflicting information for initial stop phonation type, listeners were overall more likely to base their decision on information encoded in the vowel (58% correct) rather than the consonant portion (32% correct) of the syllable. The greater influence of vocalic information held across places of articulation and vowel contexts, but not across all phonation types. However, for the stimuli in which tense and aspirated types were cross-spliced (i.e., $C*V^h$ and C^hV*), listeners' responses showed greater consonantal effects.

The perceptually most salient vowel-based information provided by syllables with original lax onsets: listeners overwhelmingly heard syllables with lax V portions as having lax onsets, regardless of the phonation type of the consonant portion. (Across consonant portions, over 4 out of 5 responses to syllables with lax V portions were judged as having lax onsets). I attribute this outcome in large part to the low F0 of lax V portions which, at vowel midpoint, averaged 35 Hz lower than the F0 of either the tense V* or aspirated V portions (see Table 1). That listeners' judgements are heavily influenced by vocalic F0 is further supported by their responses to cross-spliced syllables with V* or V^h portions: listeners rarely judged initial stops as lax if the vowel portion had a relatively high F0, again regardless of the phonation type of the consonant portion. That is, even when the consonant portions were extracted from syllables with lax onsets, with VOT values that fell between the shorter tense VOTs and longer aspirated VOTs, listeners reported hearing lax stops only 5% of the time. Thus, for these cross-spliced stimuli, low vocalic F0 was both necessary and sufficient to perceptually cue initial lax stops. This finding held for the three places of articulation and two vowel contexts tested.

When the vowel portion had a relatively high F0 (V* or V^h), listeners perceived either tense or aspirated initial stops depending on the particular combination of consonant and vowel portions. When the consonant portion was lax (i.e., with a VOT that was too long for a tense stop and typically too short for an aspirated stop), the vowel portion provided enough information to cue--although by no means perfectly--an aspirated stop for most CV^h stimuli and a tense stop for most CV* stimuli (the exceptions being [ti^h] and [pa*ŋ]). Comparison of onset F0 and H1-H2 measures of V* and V^h portions with perceptual responses to these stimuli point toward the measures' systematic influences on stop identification. For example, recall that V^h portions usually had higher onset F0 than did V* portions (Table 1). It is expected that the high proportion of ('both incorrect') aspirated responses to [pa*ŋ] stimuli is linked to the relatively small difference in F0 of vowels following tense versus aspirated labials: while labials showed a mean aspirated-tense difference of only 6 Hz at onset and overlapping aspirated-tense F0 ranges (Table 1), the vowels following velars and alveolars showed average onset differences of 22 Hz and no F0 overlap. Thus, for labials, listeners could not rely on F0 at vowel onset to differentiate tense and aspirated vowel portions. Recall also that onset amplitude measures showed positive H1-H2 differences for V^h and negative differences for V* (Table 2). Token-by-token analyses point toward listener sensitivity to these differences in low-frequency spectral tilt. For example, although the minority of identifications of [ka^hŋ] (lax consonant and aspirated vowel portions) were tense (11%), the percent tense responses tripled across the three [ka^hŋ] tokens as the H1-H2 difference decreased (from 3.3 to -4.3 dB). Similar findings hold for [t^hi] (aspirated consonant and lax vowel portions), where the token with the largest H1-H2 difference (8.7 dB) elicited two to three times as many aspirated responses as did the other two tokens. Individual token analyses showed evidence consistent with listener sensitivity to vowel duration to be sporadic at best.

Not surprisingly, token-by-token analyses also demonstrate listeners' sensitivity to acoustic differences in the consonant portion. Even though, on average, the vowel portion (V* or V^h) was the main determinant of listeners responses when the consonant portion was lax, the longer the VOT of the lax C portion, the more likely listeners were to judge the initial stop as aspirated. Although all three [pa*ŋ] stimuli were unexpectedly identified as aspirated over 50% of the time, the percent aspirated responses decreased across the three tokens as VOT duration decreased. This same pattern held for the [ta*ŋ], [ka*ŋ], and [ti*] stimuli: in all cases, the token with the longest VOT was labeled "aspirated" three to ten times more often than were either of the other tokens of that stimulus type. Moreover, although on average the [ti^h] stimuli were unexpectedly heard as ('both incorrect') tense 74% of the time, the [ti^h] token with the longest VOT was heard as ('V correct') aspirated 63% of the time.

In contrast to stimuli with a lax consonant or vowel portion, where the majority of stop identifications corresponded to the vowel portion, cross-spliced stimuli consisting only of tense and aspirated portions predominantly elicited responses that corresponded to the consonant portion. Apparently, although listeners used the V* and V^h portions to differentiate tense and aspirated stops when the VOT of the consonant portion was ambiguous with regard to the tense-aspirated distinction (i.e., CV* and CV^h stimuli), systematic VOT cues for this distinction in the consonant portion (i.e., C^hV* and $C*V^h$ stimuli) usually overrode conflicting vowel-based information. This finding is consistent with results for other languages demonstrating the particular salience of VOT (e.g., Abramson and Lisker, 1985). Once again, listeners' responses to individual tokens provides evidence of fine-grained sensitivity to VOT differences: for each place of articulation, the C^h token with the shortest VOT generally elicited the fewest (albeit still predominantly) aspirated responses. At the same time, listeners also retained sensitivity to vowel onset differences in C^hV* and $C*V^h$ stimuli. Particularly noteworthy are judgments of labial $C*V^h$ tokens (which elicited the fewest tense responses of the $C*V^h$ stimuli), where aspirated responses increased, across tokens, from 11% to 54% as the H1-H2 difference at vowel onset increased from 2.1 to 6.0 dB.

However, for the C^hV* stimuli, the main source of variability in stop identification was not the acoustic differences across tokens of a given type, or even inconsistent responses to the conflicting information in the C^h and V* portions, but rather systematic differences in individual listeners' perception of opposing tense and aspirated cues. As shown by the breakdown of responses according to individual listeners in Table 10, five listeners based their responses to one or more of the four types of C^hV* stimuli on the V* rather than the C^h portion. (Responses to $C*V^h$ stimuli did not show corresponding differences across listeners, with the exception of listener HK, whose responses across all stimulus types corresponded with the vowel portion a remarkable 92% of the time.) The striking consistency of both the consonant-based and the listeners suggests that, for the C^hV* stimuli, listeners may have been aware of the conflicting cues for initial stop phonation type and consequently chose a consonant- or vowel-based strategy for responding to these stimuli.

Table 10. Responses of each of the 12 Listeners to the C^hV* stimuli. "h" or "*" indicates that over 80% of that listener's responses to that stimulus type was C^h or C*, respectively. (The blank cell means that neither response type reached the 80% level.)

Stimulus	CP	EP	JK	JC	HK	KP	YR	SC	YY	SM	JP	YP
[pʰa*ŋ]	h	h	h	h	*	h	h	h	h	h	h	h
[kʰa*ŋ]	h	*		*	*	h	*	h	*	h	h	h
[tʰa*ŋ]	h	*	h	h	*	h	*	h	*	h	h	h
[tʰi*]	h	h	h	h	h	h	h	h	h	h	h	h

In conclusion, pulling together the results from the experiment reported here, plus previous investigations of Korean stop perception, there is now considerable evidence that neither cues provided by the vowel portion alone, nor cues provided by the consonant portion alone, are sufficient to convey the three-way laryngeal distinction across listeners, consonant place of articulation, and vowel contexts. However, for the two-way contrast between word-initial lax stops and aspirated or tense stops, low F0--phonologically, L tone--in the vowel is sufficiently and perceptually robust as to uniquely specify lax stops in the absence of any consonantal information prior to voice onset, and to override any conflicting consonantal information prior to voice onset.

References

Abramson, A. S. and Lisker, L. 1972. Voice timing in Korean stops. In *Proceedings of the VIIth International Congress of Phonetic Sciences* (A. Rigault and R. Charbonneau, editors), pp. 439-446. The Hague: Mouton.

Abramson, A. S. and Lisker, L. 1985. Relative power of cues: F0 shift versus voice timing. In *Phonetic linguistics: essays in honor of Peter Ladefoged* (V. A. Fromkin, editor), pp. 25-33. New York, NY: Academic Press.

Ahn. H.-K. 1999. *Post-release phonatory processes in English and Korean: Acoustic correlates and implications for Korean phonology.* Doctoral dissertation, The University of Texas at Austin.

Cho, T. 1996. *Vowel correlates to consonant phonation: an acoustic-perceptual study of Korean* *obstruents*. MA thesis, University of Texas, Arlington.

Cho, T. and Keating, P. 1999. Articulatory and acoustic studies of domain-initial strengthening in Korean. *UCLA Working Papers in Phonetics*, **97**, 100-138.

Han, J.-I. 1996a. *The phonetics and phonology of tense and plain consonants in Korean*. Doctoral dissertation, Cornell University, Ithaca, New York.

Han, J.-I. 1996b. Perception of Korean tense and lax consonants: evidence for a geminate analysis of tense consonants. In *Japanese/Korean Linguistics*, vol. 5 (N. Akatsuka, S. Iwasaki, and S. Strauss, editors), pp. 407-423. Stanford, CA: Center for the Study of Language and Information.

Han, M. S. and Weitzman, R. S. 1970. Acoustic features of Korean /P, T, K/, /p, t, k/ and /ph, th, kh/. *Phonetica*, **22**, 112-128.

Hardcastle, W. J. 1973. Some observations on the *tense-lax* distinction in initial stops in Korean. *Journal of Phonetics*, **1**, 263-272.

Hirose, H., Lee, C. Y., and Ushijima, T. 1974. Laryngeal control in Korean stop production. *Journal of Phonetics*, **2**, 145-152.

Jun, S.-A. 1993. *The phonetics and phonology of Korean prosody*. Doctoral dissertation, Ohio State University. Published 1996, New York: Garland.

Jun, S.-A. 1996. The influence of the microprosody on the macroprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics*, **92**, 97-116.

Jun, S.-A. 1998. The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, **15**, 189-226.

Jun, S.-A, Beckman, M. E., and Lee, H.-J. 1998. Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations for Korean obstruents. *UCLA Working Papers in Phonetics*, **96**, 43-68.

Kagaya, R. 1974. A fiberscopic and acoustic study of the Korean stops, affricates, and fricatives. *Journal of Phonetics*, **2**, 161-180.

Kim, C.-W. 1965. On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). *Word*, **21**, 339-359.

Kim, C-W. 1970. A theory of aspiration. *Phonetica*, **21**, 107-116.

Kim, M.-R. Cho 1994. *Acoustic characteristics of Korean stops and perception of English stop consonants*. Doctoral dissertation, University of Wisconsin, Madison.

Kim, M.-R. 2000a. *Segmental and tonal interactions in English and Korean: a phonetic and phonological study*. Doctoral dissertation, University of Michigan, Ann Arbor.

Kim, M.-R. 2000b. "An alternative account of so-called lax consonants." *Studies in Phonetics, Phonology, and Morphology*. Vol. 6. No.2:333-352. Hankook Munhwa Publishing co.: Seoul, Korea.

Kim, Y. 1995. Acoustic characteristics of Korean coronal stops, affricates, and fricatives. MA. Thesis. The University of Texas at Arlington.

Ladefoged, P., Maddieson, I., and Jackson, M. 1988. Investigating phonation types in different languages. In *Vocal physiology: voice production, mechanisms and functions* (O. Fujimura, editor), pp. 297-317. New York, NY: Raven Press.

Lisker, L. and Abramson, A. S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word*, **20**, 384-422.

Shimizu, K. 1996. *A cross-language study of voicing contrasts of stop consonants in Asian languages*. Tokyo: Seibido Publishing.

Silva, D. 1992. The phonetics and phonology of stop lenition in Korean. Doctoral disserta-

tion, Cornell University, Ithaca, New York.
Stevens, K. N. 1998. *Acoustic phonetics.* Cambridge, MA: MIT Press.
William M. and Beaver, R. 1994. *Introduction to Probablity and Statistics.* Duxury Press

Received: Jan. 31, 2001.
Accepted: Mar. 13, 2001.

▲ Mi-Ryoung Kim
Samsung Apt. 101-1202, Changjun-Dong
Mapo-Gu, Seoul, Korea (121-778)
Tel: +82-2-3290-1341 (O), +82-2-3143-2588 (H)
E-mail: miryoungk@yahoo.com