Prosodic structure and perception of Korean domain-initial coronal stops*

Kirk Baker
Department of Linguistics, UNC Chapel Hill, Chapel Hill, NC 27599-3155

The perception of Korean's three-way voiceless stop contrast depends on the combination of a number of acoustic properties including closure duration, f_0, VOT and spectral slope. Production studies indicate that these same properties systematically vary according to a stop's position in the hierarchical prosodic structure of an utterance. This study investigated interactions between prosodic structure and the recognition of tense, lax and aspirated stops by using stimuli that contained stops produced at one position in the prosodic hierarchy cross-spliced into utterances at a different prosodic level. The results of the experiment showed that recognition of tense and aspirated stops remained high regardless of the degree of mismatch between the underlying prosodic position of a segment and the domain into which it was inserted. Misrecognition rates were higher for cross-spliced lax stops, and underlying Utterance and Intonation Phrase initial lax stops spliced into the syllable level were perceived as aspirated by a majority of listeners. These findings are interpreted in terms of Cho & Keating's (2001) study of domain-initial strengthening and the acoustic overlap that is observed between stop manner and prosodic position. These results are taken as preliminary evidence that could support Cho & Keating's (2001) suggestion that prosodically conditioned acoustic differences aid listeners in discerning intended phrasings.

1. Introduction

Korean is well known for its typologically unique three-way voiceless stop contrast, and as a language in which prosodic structure decidedly influences the application of phonological and phonetic processes. A body of literature points to multiple spectral and temporal characteristics that distinguish the tense, lax and aspirated stops in terms of f_0, laryngeal muscle activity, glottal aperture, oral airflow rate, VOT, $H1 - H2$, intensity, closure duration, vowel duration and linguopalatal contact (see, e.g. Cho & Keating 2001, Cho, Jun & Ladefoged 2002, Kim, Beddor & Horrocks 2002 and references therein), and the role that these features play in cueing the perception of Korean consonants (Han & Weitzman 1970, Abramson & Lisker 1972, Cho Kim

* This work was supported by a Latané Summer Research Grant issued through the UNC Chapel Hill Graduate School . I am indebted to Ji Young Kim for help conducting the experiment, and to an anonymous reviewer for their helpful feedback and suggestions. All errors are mine alone.

1994, Cho 1996, Han 1996*a*, *b*, Kim et al. 2002). Production studies indicate as well that these parameters systematically vary according to the relationship between a segment and its position within the prosodic structure of an utterance (see, e.g. Jun 1993, 1995, 1996*a*, 1996*b*, Silva 1993, Jun, Beckman & Lee 1998, Ahn 1999, Kim M. R 2000, Kim S. J. 2000, Cho & Keating 2001, Cho et al. 2002).

In one such investigation into the effect of prosodic position on the realization of /n, t, t^h, t*/, Cho & Keating (2001) extend the hierarchical, intonation-based model of Korean prosodic structure of Jun (1993) and Beckman & Jun (1996) to include an Utterance level domain, adopting the 5-tiered representation shown below (from Cho & Keating 2001 p. 158).

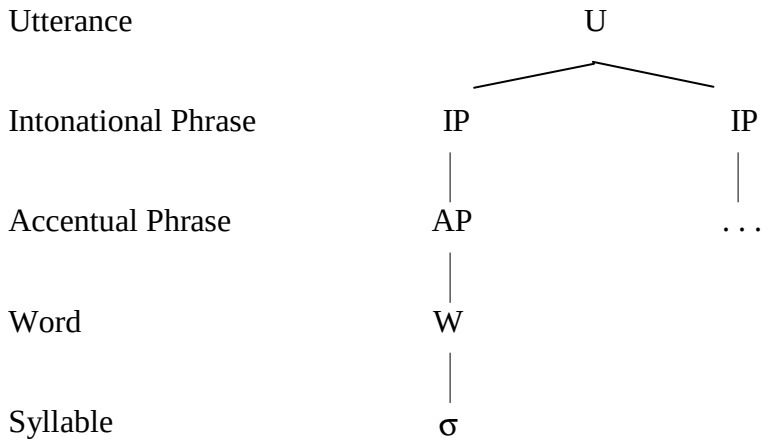

Figure 1. Korean prosodic structure (Cho & Keating 2001 p. 158 via Jun 1993 and Beckman & Jun 1996).

Cho & Keating compared a number of the articulatory and acoustic properties of domain-initial /n, t, t^h, t*/ across the five prosodic levels in Figure 1, and found that linguopalatal contact, stop duration, VOT, and total voiceless interval are greater for segments realized higher in the prosodic hierarchy; inversely, the percentage of voicing during stop closure increases for consonants realized lower in the hierarchy. They suggest that such prosodically conditioned

segmental strengthening potentially aids listeners in discerning the intonational structure intended by a speaker.

The current study addresses two points of interest raised by Cho & Keating's suggestion. It is necessary first of all to know whether the prosodically conditioned realizations of a given segment are reliably distinguished perceptually. Of corollary interest is the issue of how intonational structure interacts with the perception of the tense, lax and aspirated distinction. Specifically, this study asks whether the acoustic characteristics associated with the production of /t, t^h, t*/ for a given prosodic domain remain perceptually salient when the underlying prosodic position of the stop is altered. In order to make this determination, stops produced in one of three domain-initial prosodic contexts were excised from the original waveform and spliced into a segmentally identical context originally produced at a different level in the phrasal structure of the utterance. Based on the prosodically conditioned strengthening effects described by Cho & Keating (2001) and the overlap in acoustic parameters between stop manner and prosodic position (detailed in §2), one expectation is that the acoustic attributes of a tense or aspirated stop inserted from a higher prosodic domain into a lower one will robustly cue the original token. On the other hand, there might be tendency for a similarly cross-spliced lax stop to be perceived as aspirated. Another prediction is that the attributes of a stop inserted from a low prosodic domain into a dominant one may weakly cue the underlying token, and in some cases the manner of the underlying stop will be misperceived.

2. Perception Experiment

2.1 *Methods*

2.1.1 *Original stimuli*

The original stimuli consisted of 9 sentences in which each of the three stops /t, t^h, t*/ appeared in Utterance, Intonational Phrase, and Syllable initial prosodic position. The number of syllables

before and after the test consonants never varied (6 before, 5 after), and the immediate segmental context was always /a_ap/. These sentences are shown in Table I.

TABLE I. Utterance, IP and Syllable initial sentences for /t, t^h, t*/

Utterance initial /t, t^h, t*/ sentences						
1.	kathi mal	haypoca. /$_U$	ta	ponunke	ettay/	
	together	talk let's	all	seeing	how-Q	
	'Let's talk it over. How about seeing it all?'					
2.	kathi mal	haypoca. /$_U$	t^ha	ponunke	ettay/	
	together	talk let's	ride	trying	how-Q	
	'Let's talk it over. How about riding it?'					
3.	kathi mal	haypoca. /$_U$	t*a	ponunke	ettay/	
	together	talk let's	pick	trying	how-Q	
	'Let's talk it over. How about picking them?'					

Intonational Phrase initial /t, t^h, t*/ sentences								
4.	na	uy	chinkwu	tul a,	/$_{AP}$ ta	ponunke	ettay/	
	I	POSS	friend	PL VOC	all	seeing	how-Q	
	'Hey guys, how about seeing it all?'							
5.	na	uy	chinkwu	tul a,	/$_{AP}$ t^ha	ponunke	ettay/	
	I	POSS	friend	PL VOC	ride	trying	how-Q	
	'Hey guys, how about riding it?'							
6.	na	uy	chinkwu	tul a,	/$_{AP}$ t*a	ponunke	ettay/	
	I	POSS	friend	PL VOC	pick	trying	how-Q	
	'Hey guys, how about picking them?'							

Syllable initial /t, t^h, t*/ sentences							
7.	ayki	nun	/ocik	nata	bakke	molla/	
	baby	TOP	yet	become	except	not know	
	'Baby only knows (the word) become'.						
8.	ayki	nun	/ocik	na	t^ha	bakke	molla/
	baby	TOP	yet I	ride	except	not know	
	'Baby only knows (the phrase) "I'm riding".'						
9.	ayki	nun	/ocik	nat*a	bakke	molla/	
	baby	TOP	yet	improved	except	not know	
	'Baby only knows (the word) improved'.						

The sentences were written in Hangeul and each was read aloud 5 times by a 25 year old female native Seoul Korean speaker at a constant speaking rate and constant average f_0. She was instructed on the desired phrasings, being asked to pause after the period marking an Utterance boundary, but not to for the other two prosodic conditions. The stimuli were recorded with Kay Elemetric's CSL at an 11025 Hz sample rate using a handheld Shure SM48 unidirectional microphone in the phonetics lab at UNC Chapel Hill. Amplitude across all utterances was normalized using Kay Elemetric's ASL.

For each of the prosodic conditions and stop types, one sentence containing the domain-initial token whose acoustic parameters most closely represented the average measurements reported by Cho & Keating (2001) was selected for use in the perception study. A trained K-ToBI (Beckman & Jun 1996) transcriber checked the intonational contours of the selected sentences following criteria described by Cho & Keating (2001) and reproduced in Table II.

TABLE II. Prosodic coding criteria (Cho & Keating 2001 p. 161 via Beckman & Jun 1996)

Utterance	Period used to trigger a pause. L% boundary tone.
Intonation Phrase	Comma used to trigger IP boundary. HL% boundary tone. Final lengthening.
Syllable	Second syllable of second word in AP (LHLH) No break.

Acoustic measurements based on visual waveform and spectrogram examination of the stops in the selected sentences were made using SoundScope/16 and are given in Table III.

TABLE III. Acoustic parameters of original tokens (msecs). Due to the indeterminate silence preceding the U and IP initial stops, closure duration and total voiceless intervals were not measured.

	/t/			/t^h/			/t*/		
	U	IP	σ	U	IP	σ	U	IP	σ
V1 Duration	272	277	93	268	280	50	283	274	63
V2 Duration	54	71	68	53	54	70	96	102	66

Closure Duration			30			75			81
Voicing during closure	0	0	30	0	0	19	0	0	19
VOT	34	23	8	73	38	22	9	11	10
Total voiceless interval			38			97			91

The durations of the vowels immediately preceding and following the stops (labeled V1 and V2) were measured from onset of F2 to offset of the periodic waveform. Following Cho & Keating (2001), closure duration was measured from offset of F2 in V1 to burst beginning; voicing during closure was measured from offset of F2 in V1 to start of completely voiceless closure; VOT was measured from stop release to onset of F2 in V2; total voiceless interval is equivalent to voiceless closure plus VOT. Representative measuring judgments for all the tokens are illustrated in Figure 2 using Syllable-initial /t^h/.

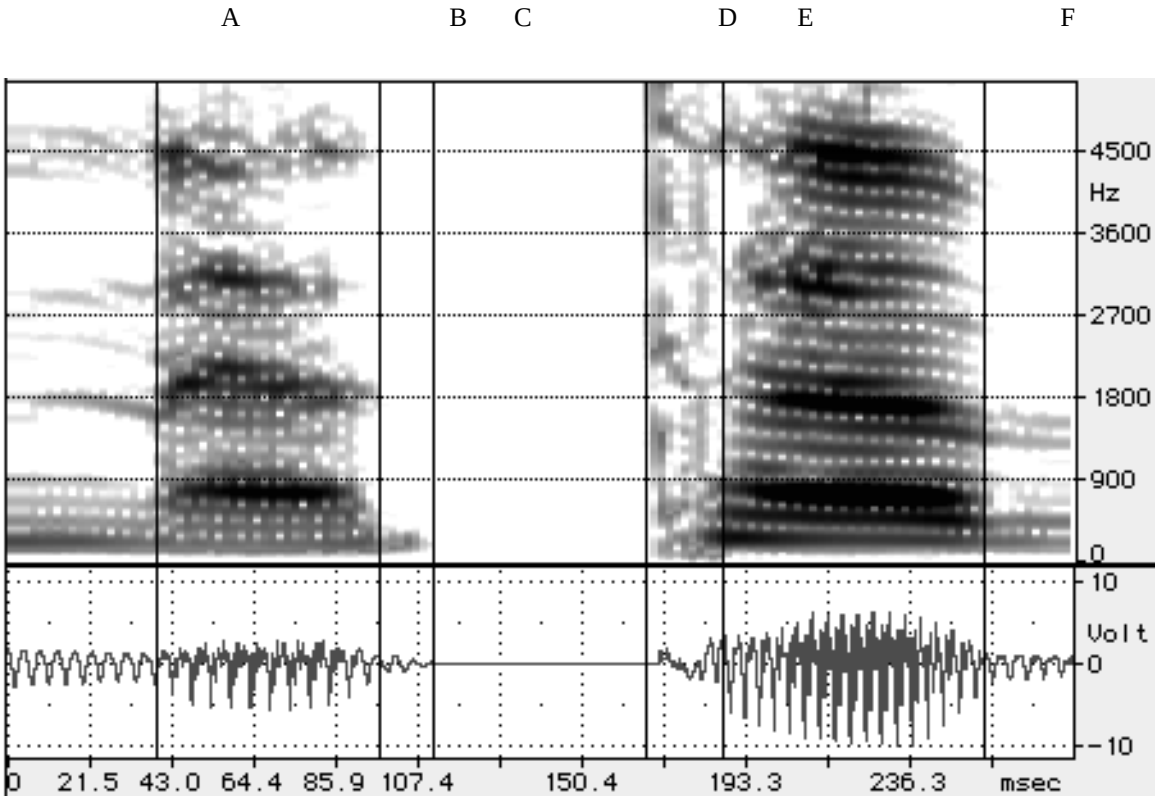

Figure 2. Waveform and spectrogram of /t^h/$_{\sigma\text{-Init}}$ illustrating measuring criteria. A – B V1; B – C: Closure Voicing; C – D: Voiceless Closure; D – E: VOT; E – F: V2

2.1.2 *Manipulated stimuli*

The stimuli used in the listening test were prepared by first copying the portion of a waveform from the onset of stop closure in the /ta, t^ha, t*a/ sequences to onset of closure of the following /p/. Each /stop vowel/ sequence was then cross-spliced into the portion of another waveform containing the identical segmental context produced at a different prosodic level. For example, to create the /ta/ conditions, original /ta/$_{\text{Utt-Init}}$ was spliced over the portions of the two waveforms containing /ta/$_{\text{IP-Init}}$ and /ta/$_{\sigma\text{-Init}}$, respectively. Next, /ta/$_{\text{IP-Init}}$ was spliced over /ta/$_{\text{Utt-Init}}$ and /ta/$_{\sigma\text{-Init}}$. Finally, /ta/$_{\sigma\text{-Init}}$ was spliced over /ta/$_{\text{Utt-Init}}$ and /ta/$_{\text{IP-Init}}$. For the /ta/ sequences that remained in their original prosodic position, the segments were cut and spliced back into the file. This procedure was followed to ensure that the splicing technique was the same for all of the stimuli. An identical procedure was repeated for /t^ha/ and /t*a/, yielding a total of 27 test conditions.

In the case of the U and IP initial stops, closure duration was inferred from the average sealed closure durations reported by Cho & Keating (2001:171), and the corresponding amount of silence was copied and cross-spliced along with V2. The inferred stop closure durations for the manipulated stimuli are shown in Table IV.

TABLE IV. U and IP initial stop closure durations for manipulated stimuli in msecs.

	/t/	/t^h/	/t*/
Utterance	80	90	100
IP	70	75	80

2.1.3 *Listening Test*

Listeners consisted of 10 female and 10 male students at Konkuk University, all native Seoul Korean speakers. The listeners were tested individually in a quiet room in the student center at Konkuk University, and all indicated normal hearing by self-report. They were told that they would hear some different sentences containing either /ta/, /t^ha/ or /t*a/ in the middle, and for each one they should point to the closest sound from predesigned answer sheets that contained these syllables handwritten in Korean orthography. The test sentences were randomized for each listener with the "Randomize" function on a Sony portable CD player (Model D-E J611) for a total of 20 scramblings of the 27 tokens. For all listeners, stimuli were delivered at a constant volume through a pair of Sony MDR-CD 180 headphones.

2.2 *Results*

Pooled responses to /t*a/ across the prosodic conditions are shown in Figure 3. Overall, 93% of responses correctly identified the underlying token in each case. For each prosodic condition, identification of the original token never fell below 90%, regardless of the direction of the splice or the distance between the original prosodic position of /t*a/ and its spliced context. However, the misidentifications that did occur were always as [t^ha] (7%).

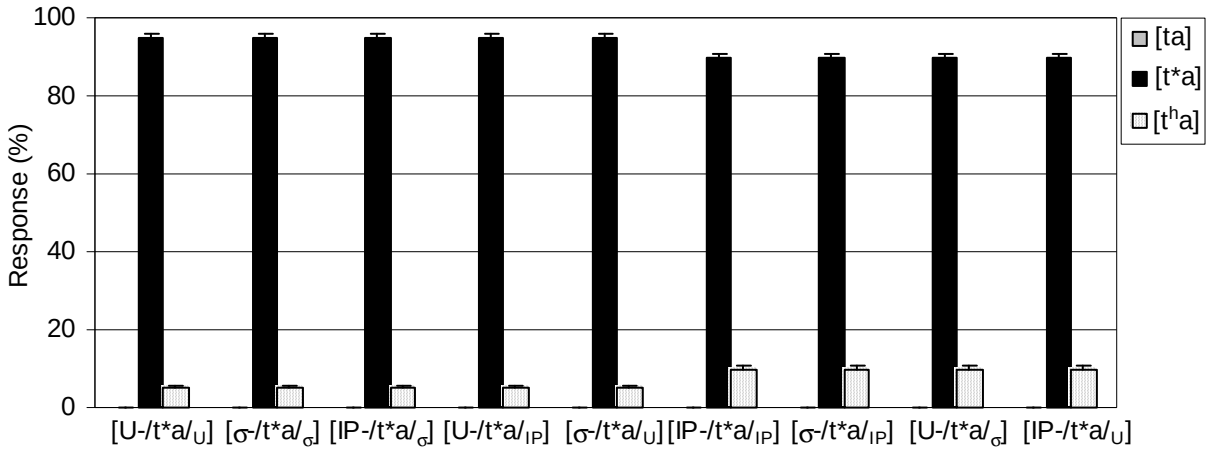

Figure 3. Pooled responses for /t*a/ across prosodic conditions. Key: [U-/t*a/$_{\sigma}$] is read as "original syllable initial /t*a/ cross-spliced into Utterance initial position", etc. Error bars represent standard error of the mean.

Similar response rates were obtained for /t^{h}a/, and are given in Figure 4. In this case, 92% of responses correctly identified the original token, and misidentifications were split between lax (5%) and tense (3%) judgments. Fisher's Exact Permutation Test shows that overall, cross-spliced conditions were more likely to be misidentified than same-spliced conditions (p=.0221). A closer look at the nature of the misidentifications reveals that /t^{h}a/$_{IP}$ was misidentified as /t*a/ at about the same rate that /t*a/ was misidentified as /t^{h}a/ (Figure 3). One subject gave [ta] responses for both the U and IP cross-spliced conditions, but the greatest number of consistent lax misidentifications occurred for cross-spliced /t^{h}a/$_{\sigma}$: inserted into IP initial position, 25% of responses were as [ta].

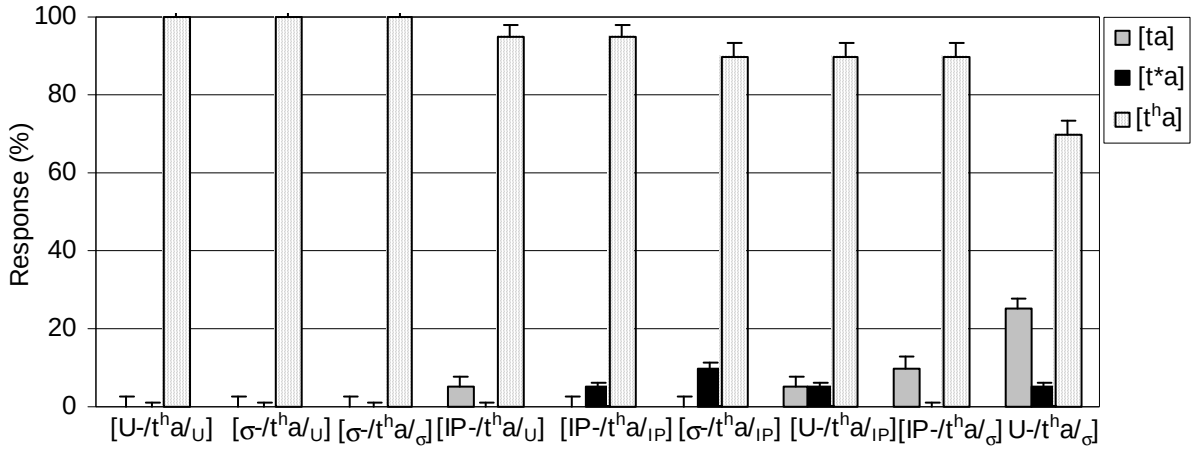

Figure 4. Pooled responses for /tʰa/ across prosodic conditions. Error bars represent standard error of the mean.

The responses for /ta/, contained in Figure 5, exhibit the most variability. Overall correct identification was 75%, and like the other stops, same-spliced and U-IP cross-spliced conditions were nearly always correctly identified. An exception to this generalization is same-spliced /ta/$_U$, for which 15% of responses were [tʰa] misidentifications; nevertheless cross-spliced conditions were more likely to be misidentified ($p<.0001$). /ta/ was misidentified as tense 7% overall, and all such misidentifications involved the syllable domain.

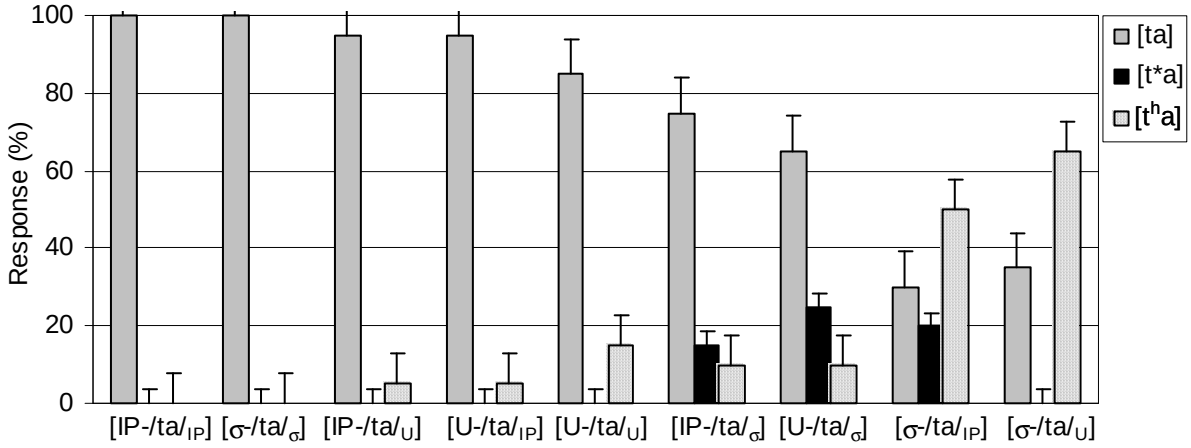

Figure 5. Pooled responses for /ta/ across prosodic conditions. Error bars represent standard error of the mean.

For /ta/$_{U}$ and /ta/$_{IP}$ inserted into syllable initial position, a majority of the responses misidentified the original tokens as aspirated. Although the limited number of subjects precludes finding a significant [t^{h}a] response tendency, 50% of listeners reported hearing [t^{h}a] for /ta/$_{IP}$ in syllable initial position, and the other responses were split between tense and plain judgments. 65% of listeners reported hearing [t^{h}a] for /ta/$_{U}$ in syllable initial position, and the remaining 35% of responses correctly identified the cross-spliced segments as lax.

3. Discussion

The fact that /t*a/ was relatively impervious to misperception may be understood partially in terms of the role that f_0 plays in cueing the lax vs. tense or aspirated distinction. In general, f_0 is lowest for vowels that come after lax consonants, and higher for vowels following tense or aspirated stops (Kim et al. 2002 p.78 and references therein). Previous perception findings demonstrated that low vocalic f_0 is sufficient to cue lax stops (Kim 1994, Han 1996a), even when paired with conflicting consonantal information (Kim et al. 2002). Given these findings, it is not surprising that none of the cross-prosodic pairings here cued the percept of [ta]. Furthermore, because /t*/ is never aspirated regardless of its position in the prosodic structure of an utterance, and because reliable aspirated percepts depend on VOT in conjunction with high f_0 and negative H1 – H2 differences (Kim et al. 2002), these stimuli would not be expected to significantly miscue [t^{h}a].

Responses to underlying /t^{h}a/ were similarly robust; the only case in which identification of the original token fell below 90% was for /t^{h}a/$_{\sigma}$ inserted into U-initial position. 25% of listeners reported [ta] percepts for this condition, which maybe expected since /t^{h}a/$_{\sigma}$ is similar to /ta/$_{U}$ in terms of closure duration, VOT, H1 – H2, and burst energy. Even in this relatively weak prosodic position, /t^{h}a/$_{\sigma}$ retained enough of its canonical aspirated status to cue aspirated percepts

for a majority of the listeners, irrespective of cross-prosodic manipulation. Additionally, the higher misidentification rate for /t^{h}a/$_{\sigma}$ in U as opposed to IP initial position could be interpreted as providing perceptual support for the Utterance level prosodic domain posited by Cho & Keating (2002) in their production study.

Underlying lax stops were most subject to misperception. Misidentifications for /ta/$_{\sigma}$ inserted into either of the two higher prosodic domains were split between tense and aspirated percepts, although overall correct identification rates remained above 65%. Once again, this pattern may be expected in the absence of VOT due to the dominance of low f_0 as a cue for [+lax] (Kim et al. 2002). Inserting an underlying higher level lax stop into syllable initial position resulted in a majority of reported [t^{h}a] percepts, although this rate never rose above 65%. The confusion in this condition follows from the fact that the strongest realization of /t/ and the weakest realization of /t^{h}/ converge in the syllable domain along the parameters of closure duration, VOT, H1 – H2, and burst energy. Although there are trends for the U-σ cross-splices to generate more misidentifications than the IP-σ conditions in both directions, the differences are not statistically significant and at best weakly attest the U-IP distinction (Cho & Keating 2002).

The overall higher misidentification rates for cross-spliced /t/ and /t^{h}/ indicate that the sub-phonemic prosodically conditioned differences in articulatory and acoustic parameters associated with stop production are perceptually distinguishable for a subset of Korean listeners. Most notably in the case of /ta/$_{U}$ and /ta/$_{IP}$ spliced into syllable initial domain, the differences are sufficient to reliably miscue the percept of the underlying token for a majority of listeners. However, it is important to note that failure to consistently misidentify the original token in terms of another does not indicate that listeners are not attending to prosodically conditioned phonetic differences. Conversely, consistent misperception of an underlying segment does not

speak directly to the issue of whether listeners utilize these prosodically based acoustic differences in discerning intended phrasing. But at the very least, the necessary condition of perceptual accessibility seems to be in place, and it seems plausible that further perceptual testing will bear out Cho & Keating's (2002) suggestion. Finally, the generally accurate identifications of the underlying stops, regardless of cross-prosodic manipulation, suggests that listeners' model of a particular segment takes into account its range of prosodically conditioned realizations. A complete understanding of Korean stop perception entails explicitly examining how prosodic structure influences the integration of the multiple acoustic cues that comprise the tense/lax/aspirated distinction.

References

Abramson, A. & L. Lisker (1972). *Voice timing in Korean stops*. In Proceedings of the VIIth international congress of phonetic sciences (A. Rigault & R. Charbonneau, eds.), pp. 439-446. The Hague: Moton.

Ahn, H. (1999). *Post-release phonatory processes in English and Korean: acoustic correlates and implications for Korean phonology*. Ph.D. dissertation, U of Texas, Austin.

Beckman, M. & Jun, S.-A. (1996). *K-ToBI (Korean ToBI) labeling conventions*. Unpublished manuscript. Ohio State University & UCLA.

Cho, T. (1996). *Vowel correlates to consonant phonation: an acoustic-perceptual study of Korean obstruents*. M.A. Thesis, U of Texas, Arlington.

Cho, T. & P. Keating (2001). *Articulatory and acoustic studies on domain-initial strengthening in Korean*. Journal of Phonetics 29(2):155-190.

Cho, T., S.-A. Jun & P. Ladefoged (2000) Acoustic and aerodynamic correlates of Korean stops and fricatives. *UCLA Working Papers in Phonetics* 99:109-148.

Han, J.-I. (1996*a*). *The phonetics and phonology of "tense" and "plain" consonants in Korean*. Ph. D. dissertation, Cornell University.

Han, J.-I. (1996*b*). *Perception of Korean tense and lax consonants: evidence for a geminate analysis of tense consonants*. In Japanese/Korean linguistics (N. Akatuka, S. Iwasaki & S. Strauss, eds.), Vol. 5, pp. 407-423. Center for Study of Language & Information, Stanford.

Han, M.-S. & R. Weitzman (1970). *Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/*. Phonetica 22: 112-128.

Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. dissertation, Ohio State University.

Jun, S.-A. (1995). *Asymmetrical prosodic effects on the laryngeal gesture in Korean*. in Phonology and Phonetic Evidence (B. Connell & A. Arvaniti, eds.), pp. 235-53. Cambridge UP.

Jun, S.-A. (1996*a*). Infl*uence of microprosody on macroprosody: a case of phrase initial strengthening*. UCLA Working Papers in Phonetics 92: 97-116.

Jun, S.-A. (1996*b*). *The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure*. New York: Garland.

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

Kim, M.-R. (2000). *Segmental and tonal interactions in English and Korean: a phonetic and phonological study*. Ph.D. dissertation, U of Michigan, Ann Arbor.

Kim, M.-R., P. Beddor & J. Horrocks. (2002). *The contribution of consonantal and vocalic information to the perception of Korean initial stops*. Journal of Phonetics 30(1): 77-100.

Kim, M.-R. Cho. (1994). *Acoustic characteristics of Korean stops and perception of English consonants*. Ph. D. dissertation, U of Wisconsin, Madison.

Kim, S. (2000). *Accentual effects on phonological rules in Korean*. Ph.D. dissertation, UNC Chapel Hill.

Silva, D. (1993). *A phonetically based analysis of (voice) and (fortis) in Korean*. in Japanese/Korean Linguistics, II. (P. Clancy. ed.), pp. 164-74. Center for Study of Language & Information, Stanford.