PERCEPTION OF TENSITY AND ASPIRATION IN SYNTHESISED KOREAN STOP CONSONANTS

John Ingram

Department of English
University of Queensland

ABSTRACT - An experiment with the synthesis of Korean Tense, Lax, and Aspirated stops is reported. Listeners responses are evaluated with identification judgements and prototypicality (goodness) measures. The status of Aspirated stops on the Tense - Lax continuum is evaluated.

INTRODUCTION

The Korean obstruents, with their three way contrast between Tense, Lax, and Aspirated stops and affricates ([p'aŋ], *bread*; [paŋ], *room*; [p^{h}aŋ], *bang*) provide an unusual example of the independent control of the features of Tensity and Aspiration (Kim, 1965). While they tend to cluster in distinct regions of the Voice Onset Time (VOT) continuum, differences between the three types of Korean stops cannot be accounted for solely in terms of the timing of the oral release in relation to the voicing of the accompanying vowel. Differences in laryngeal and supra-laryngeal setting, traditionally characterised in terms of a tense-lax distinction, need also to be considered, forming a two-dimensional space of phonetic/phonological contrasts. English learners of Korean have particular difficulty hearing and producing this 3-way contrast, though it is eventually mastered.

While there is general agreement about the dimensionality of the phenomenon, there remains disagreement over the location of the Aspirated Korean obstruents within this two dimensional phonetic space (Kim, 1965; Abramson and Lisker, 1971; Hardcastle, 1973). Aspirated stops have been regarded as "tense" by some, and as "lax" by others. Three hypotheses for the status of the aspirated stop on the tense-lax continuum are illustrated below.

```
             Model 1              Model 2      Model 3
          Tense   Lax             T     L      T     L
Unasp.    p'----- p  Short        p' ---- p    p'----- p
          |          VOT                  |          |
          |                               |          |
Asp.      ph         Long                 ph         ph
                     VOT
```

Fig. 1. Three hypotheses for the status of Aspirated stops

The first aim of this study was to resolve the phonetic status of Korean Aspirated stops, using speech synthesis and perceptual ratings by native Korean listeners. It was proposed to construct a two dimensional stimulus array of tokens, ranging over phonetically plausible values of Aspiration (voicing) and Tensity (laxness), and then to use native-listener identifications and goodness ratings of the stimuli, to select perceptual prototypes and thereby determine phonetic relations among the stimuli, and in particular, to determine the location of the Aspirated stops with respect to the Tensity dimension.

The larger aim of the project is to use speech synthesis to investigate the role of phonological experience in the perception of phonetic features. Presumably English learners of Korean need to adapt a uni-dimensional aspiration/voicing contrast to a two-dimensional set of phonemic oppositions in acquiring Korean stop contrasts. Speech synthesis provides a potentially powerful tool for investigating perceptual changes in second language learning. To the author's knowledge, there has been only one previous attempt to synthesise Korean Tense, Lax and Aspirated stops (Lee, 1990). If the synthesis proved adequate for resolving the status of Aspirated stops in Korean then it could be extended to investigate *perceptual magnet* effects (Iverson & Kuhl, 1995) in native listeners and second language learners of Korean. Preliminary findings from this extension of the study will be reported in the oral presentation.

PREVIOUS FINDINGS: KOREAN STOPS

Hardcastle (1973) characterised the Tensity feature of Korean stops in terms of heightened isometric muscular tension in the vocal folds and the pharynx. Subsequently, this view has been supported by acoustic and physiological observations of tense and lax Korean stops and affricates. Acoustically, a tense supralaryngeal muscular setting may be expected to express itself in a lower degree of compliance of the vocal tract walls, yielding sharper vocal tract resonances and less damping of higher harmonics of the voice source. Increased stiffness in the laryngeal muscles may be expected to yield a glottal tone with a steeper spectral slope and a higher fundamental frequency at voice onset following the stop release. A more fortis articulatory setting would also be expected to produce stops with abrupt oral release and vowel onset characteristics, evident in waveform or intensity traces.

These acoustic features have been confirmed by acoustic comparisons of Korean Tense and Lax stops (Kim, 1965; Han and Weitzman, 1970). They have also been supported by physiological observations. Using electromyographic recordings from the intrinsic muscles of the larynx using hook wire electrodes, Hirose, et al. (1974) found that a brief burst of vocalis muscle activity occurred just prior to the oral release in the Tense stops. That is, a brief (100-200 msec) "stiffening" of the vocal folds was found in the closed phase of Tense stops, which was absent in the Lax and the Aspirated stops. Aspirated stops showed a sharp burst of activity in the *interarytenoid* muscles, supported by activity of the *vocalis* muscle, during the *release* phase, not found in the tense or lax stops. As well as abducting the vocal folds, this gesture would increase the stiffness of the folds, helping to suppress vocal fold vibration during the aspiration phase. Thus, electromyographic observations of intrinsic laryngeal muscles suggest increased tension in the vocal folds is a characteristic of both the Tense and the Aspirated stops, though the phasing of the gesture with respect to the oral release is different in each case. However, on other indices of Tensity, the Korean aspirated stops have been found to cluster with the Lax stops (see below).

THE SYNTHESIS EXPERIMENT

Natural tokens

Five native speakers of Korean provided tokens of Korean stops from word list readings recorded in a studio at the University of Queensland. These natural tokens provided parametric values for synthesis and were saved for future use in perceptual evaluations of the synthetic stimuli. Various measurements and qualitative observations were made of the aspiration bursts, formant frequencies, f0 values - particularly around vowel onset, which were relevant for synthesis of stop consonant series. In agreement with previous studies, a relatively high f0 in the vowel onset following Aspirated stops was noted, comparable to the Tense stops. But otherwise, waveform and spectral characteristics of the Aspirated stops more closely resembled the Lax than the Tense stops.

Synthesis method

Three series of stop consonants, one for each of the labial, alveolar, and velar places of articulation were synthesised on a Klatt cascade/parallel formant synthesiser. Each stimulus array covered the perimeter of a two-dimensional phonetic space based on, 1) the duration of Aspiration (Voice Onset Time: VOT), and 2) Tensity, as encoded by acoustic parameters reflecting the stiffness of the vocal folds, the compliance of vocal tract walls, and the effort with which a plosive is articulated. The Klatt parameters used for control of the Tensity dimension were:

1. Higher f0 in the first few glottal cycles of the following vowel, caused by increased stiffness of the vocal folds. KLSYN parameter: f0
2. Formant bandwidths in the vicinity of the vowel onset: narrower for Tense articulatory setting. KLYSN parameters: B1, B2, B3
3. The open quotient of the glottal volume velocity waveform. The open quotient is higher with slack vocal folds (lax glottal setting). KLSYN parameter: OQ
4. The amplitude of the vowel at onset. Tense setting associated with more abrupt and higher amplitude vowel at onset. KLYSN parameter: AV

Eleven synthetic tokens formed each stimulus array. Five of these stimuli were modelled closely on natural tokens and have been labelled "prototypes" or "potential perceptual magnets". Six other stimuli, represent intermediate targets, intended to fall in regions of phonetic space, between phonemic targets. The synthetic stimulus array was constructed in the following way.

(i) Two short-lag VOT tokens were synthesised, representing a prototypical Tense and a Lax initial stop on a _VC carrier syllable. Klatt parameter settings for these initial synthetic tokens were derived, where possible, from measurements taken from the natural tokens of one of the native speakers. Where parametric values could not be measured, initial values were determined by trial and error, through auditory comparisons of synthetic and natural tokens. Tensity parameter settings for the short-lag Tense and Lax prototypes are shown in Table 1.

KLATT PARAMETER	TENSE STOP	LAX STOP
f0 at vowel onset:	140 Hz	111 Hz
AV at vowel onset:	60 dB	50->60 dB over first 40 msec.
OQ at vowel onset:	50% fixed	90->50% over first 40 msec.
B1 - B4:	default, fixed	twice default -> default over first 20 msec.

TABLE 1. Klatt Tensity synthesis parameters for tense and lax stops

(ii) The two short-lag Tense and Lax synthetic tokens were modified to produce long-lag (aspirated) variants forming the lower corners of the stimulus rectangle as shown in Fig. 2. below, where the "x's" represent prototypes or potential perceptual magnets. The aspiration bursts for these long-lag stimuli were modelled on the spectral characteristics of natural Aspirated tokens, using the Klatt parameters AI, AF and AH. This produced "tense" and "lax" Aspirated synthetic tokens.

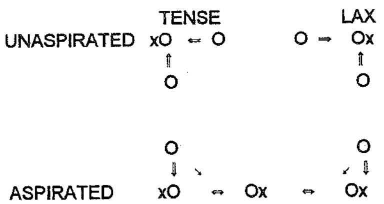

Fig. 2. Synthetic stimulus array used in the study. The "O"s mark the 11 stimuli generated. The "x's" mark potential magnets. Arrows indicate the direction of predicted magnet effects.

(iii) A fifth "potential magnet", a synthetic Aspirated token of "mixed" tensity was constructed, using a "tense" high f0 value at vowel onset in combination with otherwise "lax" parameter values for the Klatt Tensity feature. To the experimenter's ear, and in spectrographic appearance, this "mixed" Aspirated prototype was closer to the natural Aspirated Korean target than were either of the "tense" or "lax" synthetic Aspirated prototypes.

(iv) The 6 non-prototype tokens were generated using intermediate parametric values defined in relation to their nearest prototypes. Parameter settings that were 25% up or down from those of the nearest prototype were used. This rule of thumb produced a VOT change of 10-15 msec.

PERCEPTUAL EXPERIMENT

Subjects and method

Five male adult native speakers of Korean, two of whom were fluent in English and three of whom spoke very little English, undertook identification tests and rated each synthetic token on 7 point "goodness" scale, to provide a measure of stimulus prototypicality. The synthetic Labial, Alveolar and Velar stops were presented in three separate blocks of 22 randomised trials. There was a repetition of each token in the 11 item stimulus set. On a given trial, each stimulus was presented three times with an ISI of .75 sec. over headphones.

Results

Table 2 summarises the results of the Identification tests and the Prototypicality ratings for the 3 sets of synthetic stops.

Rates of identification for the short-lag Tense and Lax stops were uniformly high across the three places of articulations (i.e.: correct hits on 10/10 trials). The Aspirated stops were not as well identified, particularly for the labial and velar series, where the highest rates of identification were 7/10 and 5/10 respectively. Only for the alveolars were hit rates of 10/10 achieved, for the "mixed" and the "tense" Aspirated prototypes. Overall, the identification test results favoured the "mixed" tensity Aspirated stops over the "lax" and "tense" prototypes.

The "goodness" or prototypicality ratings for particular stimuli was highly correlated with their identification scores. Stimulus tokens that were designed during synthesis as prototypes generally obtained higher goodness ratings than those synthesised with "intermediate" parametric settings of Aspiration and Tensity (stimuli with numerical designations 1-6 in Table 2). The "mixed" tensity Aspirated stops obtained higher "goodness" ratings than the "tense" or the "lax" Aspirated tokens.

Conclusions

Control of the Tensity feature proved to be more successful than the Aspiration feature in the synthesis of Korean stops. It was difficult to produce sufficiently high amplitude, low frequency aspiration noise to satisfactorily model the Korean labial and velar Aspirated stops. Further work is needed to establish whether or not this limitation is intrinsic to the Klatt synthesiser.

The "mixed" Tensity setting yielded the highest identification scores and prototypicality ratings for Aspirated stops. The higher f0 at vowel onset is probably attributable to increased stiffness in the vocalis muscle as the glottis is rapidly closed from a wide aperture during the aspiration phase. However, aspirated stops lack the generalised muscle tension and increased pulmonic effort that accompanies Tense stops in Korean.

REFERENCES

Abramson, A.S. and Lisker, L. (1971) *Voice timing in Korean stops.* Status report on Speech Research, no. 27 Haskins Laboratories: 179-185.

Han, M.S. & Weitzman (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. & Ushijima (1974) *Laryngeal control in Korean stop production.* **2**, 145-152.

Iverson, P & Kuhl, P.K. (1995) *Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling.* J. Acoust. Soc. Am. **97**, 553-562.

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

Lee, S.-H. (1990) *Korean lenis and fortis stops: synthesis and categorical speech perception task.* Working papers in Linguistics, Columbus Ohio, **38** 105-120.

Identification responses					Prototypicality ratings			
LAX			TENSE					
p	p1	P2	P		p	p1	P2	P
p P p^h	p P p^h	p P p^h	p P p^h		**5.7**	4.7	6.2	**6.8**
10 - -	1 9 -	- 10 -	- 10 -					
p6			P3		p6			P3
p P pH			p P pH		5.3			5.6
9 - 1			- 10 -					
p^h5			P^h4		p^h5			P^h4
p P p^h			p P p^h		3.7			4.6
7 2 1			- 8 2					
p^h	p^hf0		P^h		p^h	p^hf0		P^h
p P p^h	p P p^h		p P p^h		4.3	**5.4**		4.7
5 - 5	- 3 7		- 7 3					
	ASPIRATED							
k	k1	K2	K		k	k1	K2	K
k K k^h	k K k^h	k K k^h	k K k^h		4.6	5.2	**6.9**	5.0
10 - -	1 9 -	- 10 -	- 10 -					
k6			K3		k6			K3
k K k^h			k K k^h		**4.9**			6.4
10 - -			- 10 -					
k^h5			K^h4		k^h5			K^h4
k K k^h			k K k^h		4.8			5.0
10 - -			- 10 -					
k^h	k^hf0		K^h		k^h	k^hf0		K^h
k K k^h	k K k^h		k K k^h		3.7	**4.9**		4.5
9 - 1	- 5 5		- 7 3					
t	t1	T2	T		t	t1	T2	T
t T t^h	t T t^h	t T t^h	t T t^h		**6.8**	4.5	6.7	**6.8**
10 - -	9 1 -	- 10 -	- 10 -					
t6			T3		t6			T3
t T t^h			t T t^h		5.2			5.0
10 - -			- 10 -					
t^h5			T^h4		t^h5			T^h4
t T t^h			t T t^h		2.3			3.8
4 - 6			- 7 3					
t^h	t^hf0		T^h		t^h	t^hf0		T^h
t T t^h	t T t^h		t T t^h		3.9	**5.7**		4.8
1 - 9	- - 10		- - 10				mean scores	
		raw frequencies						

Table 2. Identification responses and Prototypicality ratings for synthesised Korean stops