Speaker-specific Implementation of VOT Values in Korean*

Jeong-Im Han** · Jooyeon Kim***

ABSTRACT

The purpose of the present study is to test whether VOT values of the Korean plain stops in intervocalic position are encoded differently by individual speakers. In Scobbie (2006), the VOT values to the /p/-/b/ voicing contrast in Shetland Isles English were found to demonstrate a high degree of inter-speaker variation. More importantly such variation was not arbitrary: first, there was an inverse relationship between the amount of prevoicing for /b/ and the duration of aspiration for /p/. Second, the inter-speaker variation was shown to be similar between the subjects and their parents. These results suggest that the phonetic targets for VOT are specified in fine detail by speakers. The present study further explores this issue in terms of testing 1) whether the likelihood and the amount of voicing for the intervocalic plain stops in Korean show inter-speaker variation; 2) whether the likelihood and the exact amount of voicing for the intervocalic plain stops in Korean are closely related to the amount of aspiration for the Korean intervocalic aspirated stops. The results of the study suggest that the voicing of intervocalic plain stops in Korean varied according to the individual speakers, but it did not seem to be directly interrelated with the amount of aspiration of the aspirated stops in the same phonological position.

Keywords: interspeaker variation, VOT, Korean intervocalic stops, exemplar models of mental lexicon

1. Introduction

1.1 Speaker-specific sound systems

In a traditional approach espoused by Chomsky and Halle (1968, SPE), phonology and phonetics belong to distinct modules: phonology is the domain of abstract patterns understood to be discrete and categorical, while phonetics is the domain of the quantitative realization of those patterns, which is outside the grammar. A great number of studies have been put forth based on this approach, accounting for many phonological processes in world languages. However, in such strict modular view of grammar, many processes which show both

* The present research was supported by Konkuk University in 2008.
** Department of English Language and Literature, Konkuk University
*** Department of Korean Language and Literature, Konkuk University

phonological and phonetic properties remain unexplained. For example, the vowel before a voiced consonant as in *bead* is generally longer than that before a voiceless consonant as in *beat*. However, such difference is greater in English than many other languages (Keating, 1985). Similarly, final stops in Thai are always unreleased (Abramson, 1972), while those in German are almost always released (Martens & Martens, 1965), which is unlikely to be derived from other aspects of the language, and thus must be specified. These language-specific differences cannot be accounted for in a modular approach. Another problem with the modular approach is that it cannot capture the functional similarity in many processes occurring in phonology and phonetics at the same time. For example, assimilation is a phonological process, but coarticulation is considered as a purely phonetic implementation, even though the functional principles triggering these two phenomena are the same (Flemming, 1997).

More recently, exemplar theories of phonology call into the question this modular approach to phonology and phonetics (Johnson, 1997, 2007; Bybee, 2001a, b; Pierrehumbert, 2001, 2002, 2003; Pierrehumbert et al., 2000; Goldinger, 1997; Mullenix, 1997; Pisoni, 1997; Hawkins & Smith, 2001; Hawkins, 2003; Foulkes & Docherty, 2006). These studies are based on the results of psycholinguistic researches on mental lexicon that unlike the generative approach, people are not born with basic grammar of phonology, but they create their own phonological and phonetic systems through statistical patterning in the input and universal linguistic and cognitive predispositions. Given the assumption that people perceive phonetic details subtle enough to convey both a lexical contrast and linguistically non-contrastive indexical factors, lexicon is a mixture of abstract knowledge and detailed memories of previous speech events. In this approach the language sound systems are represented in the set of phonetically detailed exemplars of speech that people experienced, and also phonological rules or phonological entities like phonemes emerge from the detailed exemplars.

Exemplar models are of particular interest to many phoneticians who try to show socially conditioned phonetic variation (Thomas, 2002; Labov, 1994; Foulkes & Docherty, 2006). Much research in this framework has shown how fine-grained, but consistent pattern of phonological and phonetic variation appears within a single language or dialect as well as cross-linguistically. Foulkes and Docherty (2006) provided a wide range of examples to show that phonetic variants are associated with particular speaker groups or particular speaker styles, and that even children as well as adults show such variation at the first stages of acquisition. Harrington (2006) presented a longitudinal study of some vowels from the annual Christmas broadcasts produced by Queen Elizabeth II over a 50 year period to test whether adults adapt to sound changes taking place in the community. It was shown that Queen's final vowel in *happy* was less tense in 1950s data than that in 1980s data, suggesting that it has undergone phonetic raising in Received Pronunciation (RP).

Recently Scobbie (2006) further showed a case for the socially conditioned phonetic variation within a speaker. He examined the VOT patterns in a single speech community, the Shetland Isles of Scotland, which are an isolated archipelago of about 15 inhabited and more than 80 uninhabited islands, west of Norway. He recruited young adult Shetlanders whose parents were also native Shetlanders as well as Shetlanders whose parents were born and raised in Scotland and England, and then measured the VOT values of the two phonemes, /p/ and /b/. Even though the /p/ and /b/ contrast was salient phonemically in Shetlandic English, individual speakers encoded their own VOT targets. Many speakers whose parents had been born and raised in the Shetland Isles seemed to have a strict categorical opposition of prevoiced [b] and short lag [p], while speakers whose parents had been from England tended to have short lag [p] and long lag [p^h]. However, it is noteworthy from the individual speaker results that the more prevoicing for [b] that a speaker used, the shorter their VOT for [p] could be without causing perceptual confusion. Thus there was generally an inverse relationship between the rate of prevoicing of /p/ and the mean VOT of /p^h/. These results indicate that phonetic targets for VOT are specified in fine detail by speakers, suggesting that indexical and phonological cues are simultaneously present.

1.2 VOT variation in Korean

It is widely assumed that voiceless plain stops are voiced between two voiced segments in Korean, and this process is argued to be categorical modification of the target consonants in most earlier analyses based on a phonological approach (Cho, 1987, 1990; Kim-Renaud, 1974; Kang, 1992). However, work based on phonetic experiments addresses a fundamental question with respect to such categorical explanation of the plain stop voicing process, since the voicing process shows variation depending on the context. According to Jun (1995), a plain stop is more likely to be voiced in faster rate, in shorter duration, and next to a segment with a stronger voicing gesture. Further voicing is sensitive to the prosodic domains. The results of the acoustic measurements (Silva, 1992; Jun, 1995) show that plain stops are voiced within words, but are progressively less voiced in the edge of the bigger prosodic categories. Based on these evidence, Jun (1995) argues that "Lenis [plain] Stop Voicing in Korean is not a phonological rule. Rather ... a gradient process" (p. 248). Recently, some phonetic work (Han, 2000) showed that even in word-internal structure, the voicing distinction in Korean is not categorical, and thus not congruent with the phonological analyses considering voiced stops as categorical outputs.

1.3 Proposal

Based on the findings from a very fine within-language variation, the present study examines 1) whether the variation of VOT in word-internal intervocalic plain stops is speaker-specific, and 2) the VOT variation word-internal intervocalic plain stops is closely related to that of aspirated stops in the same phonological/prosodic position. Specifically, the present study measures the extent of voicing in two different dimensions: 'the amount of voicing', the percentage of duration which is voiced as compared to the whole closure duration, and 'the likelihood of voicing' which refers to how often a fully voiced closure appears (Smith, 1997; Han, 2000). These two dimensions are investigated, examining whether there is any speaker variation in values of these two dimensions. Based on these results, the subjects are grouped according to the pattern of the implementation of VOT values. This study also examines the VOT values of the aspirated stops in word-internal intervocalic position to see any consistent pattern between the amount of prevoicing of plain stops and the duration of aspiration for aspirated stops. If the plain stops are produced with full voicing, they are easily distinguished from the corresponding aspirated stops. However, if they are produced with a small amount of voicing, or even as voiceless consonants with a positive value of VOT, they are not easily differentiated from the corresponding aspirated stops, and thus the aspirated stops should be produced with much large amount of VOT. The results of the experiment reported here thus could provide further support for the claim by Scobbie (2006) that there is flexibility in the implementation of the VOT system.

2. Method

2.1 Stimuli

The test words consisting of V_1CV_2 sequences were constructed, in which the V segments (V_1 and V_2) were /i/, /ɛ/, /ɨ/, /u/, /o/, /ɔ/, or /a/, and C segment was the bilabial stop /p/. The reason for the place of articulation for stops limited to a bilabial position is that this position is more clearly displayed and thus more easily discernable, as compared to other places of articulation (Silva, 1992). The selection of the vowels is based on the assumption that modern Korean has seven vowel phonemes (Shin, 2000). Each test word was recorded in the frame sentence as in *ikɔ irɨmi* ______ *ta* ('The name of this is ______').

2.2 Subjects and Recording

Seven male speakers, who were college students, participated in the recording. They were native speakers of the standard Korean and had no reported history of either speaking or hearing disorders. They were all born and raised in Seoul, Korea. They were recorded in a sound-proofed booth, using a Shure SMZ headworn microphone and Marantz PMD 670 solid-state cassette recorder. Before recording, each speaker was asked to read a few randomly chosen test sentences to familiarize themselves with the materials. Then they were asked to read each test utterance at a natural, comfortable speed. The test sentences were presented in lists in Korean orthography and the order of them was randomized.

2.3 Analysis

The recorded data were digitized at a sampling rate of 22,050 sample/second, and then stored as files to be processed by PRAAT. Waveforms, and spectrograms for each token were created. The total number of tokens was 587 (49 words x 4 speakers x 3 repetitions). The marks were placed on the waveforms as well as spectrograms. When the two displays appeared to show minor discrepancies, the mark was placed with reference to the waveform. The cues for the beginning of stop closure were the reduction of complex, periodic waves in the waveform and the attenuation of the second formant structure of the preceding vowel in the spectrogram. The beginning of stop release was captured with a sudden increase in energy in both waveform and spectrogram. The amount of voicing during stop closure was measured with periodic vibration in the waveform and a voice bar in the spectrogram during the interval of stop closure. Again if there are minor discrepancies between the displays of waveforms and spectrograms, the former was the main cue. The cue for the beginning of the vowel onset was the beginning of the second formant structure.

3. Results

The results of 'the likelihood of voicing' which refers to how often speakers pronounce the target consonants as fully voiced are presented in <Table 1>. For each utterance containing a target stop, the likelihood was calculated for each speaker by counting the number of repetitions of those utterances that were fully voiced during the whole interval of closure.

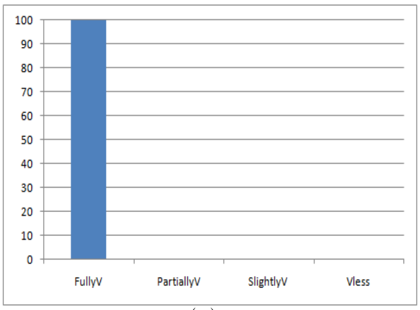
<Table 1> indicates that only one speaker (speaker 1) showed full voicing in every token. Speaker 7 showed only small number of tokens which were fully voiced (4.1%). Overall intervocalic plain stops were realized with a varying degree of voicing depending on speakers. The mean percentage of voicing (42.8%) shows that intervocalic voicing of the plain stops is not a categorical change as proposed in phonological analyses.

Table 1. The overall percentage of fully voiced tokens produced by each speaker and the whole duration of the target words (ms)

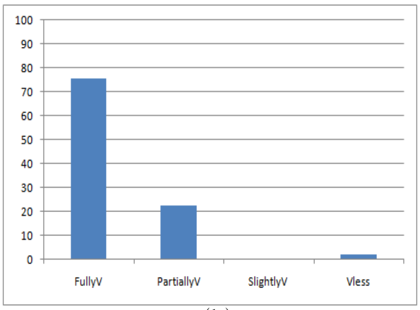
speaker	percentage of voicing	total number of tokens	tokens which were voiced	duration of the target words (ms)
speaker 1	100%	49	11	205
speaker 2	75.5%	49	37	245
speaker 3	63.8%	49	2	308
speaker 4	22.4 %	47	6	278
speaker 5	21.3%	47	10	253
speaker 6	12.8%	47	30	222
speaker 7	4.1%	49	49	285
mean	42.8%			

Now turn to the specific pattern of voicing for each speaker. For convenience, 'the amount of voicing', the percentage of duration which is voiced as compared to the whole closure duration, is represented as the following four categories: 'fully voiced tokens' (100%), 'partially voiced tokens' (41-99%), 'slightly voiced tokens' (1-40%), and 'voiceless tokens' (0%). The percentage of each category for each speaker is as in <Figure 1>.

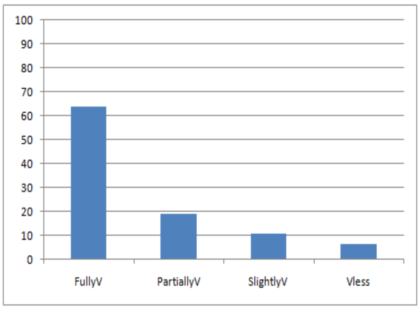
<Figure 1> clearly shows that speakers varied as to how many of their tokens fell into each of the four voicing categories. Speaker 1 produced all tokens as fully voiced; Speaker 7 produced most of tokens as devoiced; All the other speakers exhibited various patterns of voicing in between, showing various VOT values for the four categories.



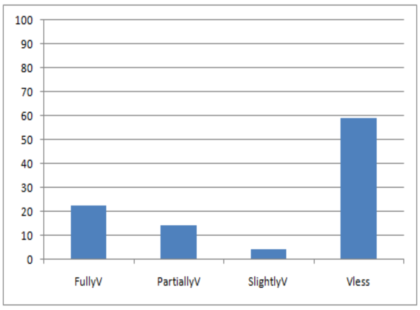
(a)



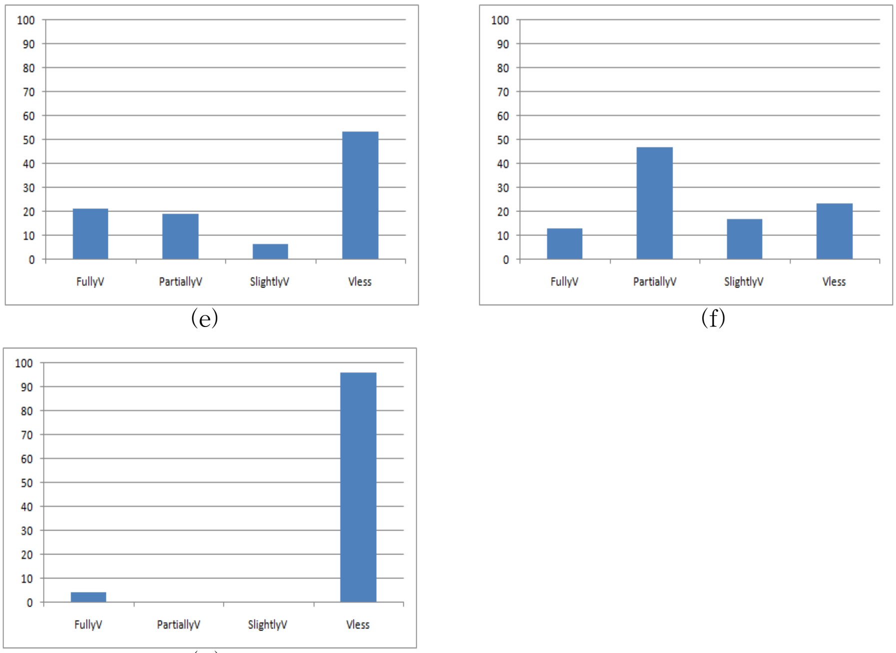
(b)



(c)



(d)



(e) (f)

(g)

Figure 1. The mean percentage of number of tokens that were fully voiced, partially voiced, slightly voiced, and voiceless for each speaker: (a) for Speaker 1, (b) for Speaker 2, (c) for Speaker 3, (d) for Speaker 4, (e) for Speaker 5, (f) for Speaker 6, and (g) for Speaker 7 (fully voiced=100% of voicing, partially voiced=41-99% of voicing, slightly voiced=1-40% of voicing, voiceless=0% of voicing).

Given that the plain stops which were produced as devoiced in intervocalic position can cause perceptual confusion with the corresponding voiceless aspirated stops, it is worth to see whether there is any systematic relationship between the likelihood of voicing and the mean VOT values of the aspirated stop, /p^h/. It is expected that to maintain contrast of the plain stops and aspirated stops in intervocalic position, there should be inverse relationship between the number of fully voiced tokens and the mean VOT values of the aspirated stops.

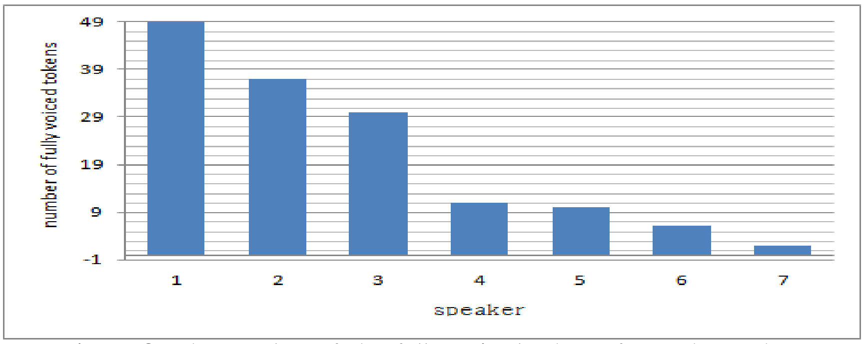


Figure 2. The number of the fully voiced tokens for each speaker.

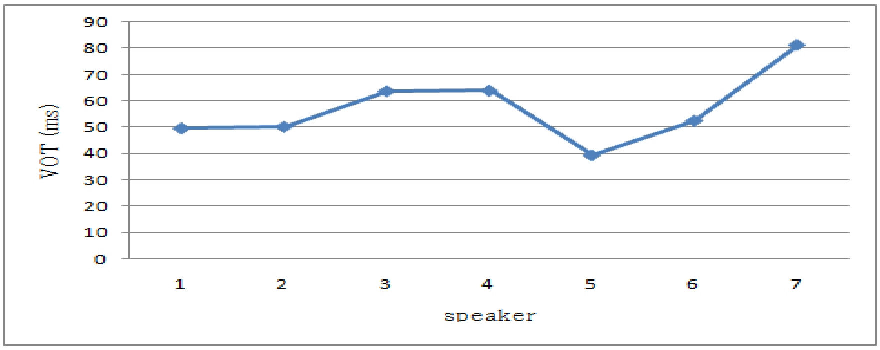


Figure 3. The mean VOT values (ms) of the aspirated stops in intervocalic position for each speaker.

The comparison of the results between <Figure 2> and <Figure 3> reveals that the speakers whose numbers of fully voiced tokens were really small were likely to produce the aspirated stops with large amount of VOT (Speaker 7), but there was no direct inverse relationship between the rate of full voicing and the mean VOT of the aspirated stops.

4. Discussion

It was expected from Scobbie (2006) that 1) the /p/ and /p^h/ contrast in intervocalic position seems to be shared at some abstract level, but individual speakers encode their own VOT targets; 2) the VOT variation is not random but systematic in that speakers who produce the plain stops in this position as voiceless (devoiced) are likely to show longer aspiration for the corresponding aspirated stops to preserve the contrast between these two types of consonants.

First of all, the results of the present study revealed that even in word-internal, intervocalic position, speakers did not show any consistent pattern of VOT values. Speakers varied considerably as to not only the number of tokens that were fully voiced, but also the amount of voicing during the stop closure. The gradient increase in mean VOT for the plain stops across the subjects means that there is no basis for allocating /p/ into a voiced stop category. These results suggest that voicing in intervocalic position might be a categorical change at some abstract level as assumed in phonological analyses but, there is no consistent pattern to represent the distinctive feature of [+voice] and thus individual speakers encode their own VOT targets.

However, it is not clear whether there is an inverse relationship between the rate of full voicing of /p/ and the mean VOT values of /p^h/ in word-internal, intervocalic position. One speaker (Speaker 7) who produced almost all tokens of intervocalic plain stops as voiceless showed relatively large amount of aspiration when he produced the aspirated stops in the same phonological position. However, all the other speakers did not show such clear tendency.

The discrepancy in the results between Scobbie (2006) and the present study might be explained with the fact that the inter-speaker variation was displayed in different phonological relations between Scobbie (2006) and the present study. Namely, [p] and [b] in the Shetlandic English and the Scottish English were associated with separate phonemes, while [p^h] and [b] in Korean were not. The aspirated stop in Korean (/p^h/) is a separated phoneme, but the voiced plain stop, [b], appears as an allophone of /p/. Namely, the voicing of the plain stops in intervocalic position does not play a role in the phonemic contrast. This might lead to the result that though devoiced plain stops and the aspirated stops are acoustically similar, the impact of this similarity might be really small.

Alternatively, Korean speakers might rely more on the vocalic cues than the consonantal cues to identify the consonants. They might have some difficulties in discriminating the plain and aspirated stops in intervocalic position with similar VOT values, but they could get additional information to identify those stops from the acoustic cues of the following vowel. Kim et al. (2002) investigated the relative importance of the consonantal and vocalic information to the perception of the Korean word-initial stops. They found that for the two-way contrast between word-initial plain stops and aspirated stops, low F0 in the vowel (vocalic information) is sufficiently perceptually robust as to uniquely specify plain stops in the absence of any consonantal information such as voice onset time, and to override any conflicting consonantal information prior to voice onset time. These results might have to do with the gradient merger of VOT values between the plain stops and aspirated stops. Silva (2006) suggests that the difference in VOT between phrase-initial stops of the aspirated stops and the plain stops is reducing now to the point of merger among younger speakers, but at the same time the contrast between the two stops is being maintained through otherwise redundant pitch differences in the following vowel. Given that VOT was not a robust indicator to the contrast for plain stops and aspirated stops, the VOT values of the two stop categories might be overlapped without any risk of losing contrast.

It is not clear whether the results of the present study support the exemplar theory of mental lexicon. Within a modular approach, a categorical analysis reveals the simple structure underlying variable data, not showing the detailed linguistic distinctions and patterns of variation that the speaker make. An exemplar model, on the other hand, predicts flexibility among speakers and permits flexibility in the linguistic analysis between gradient and categorical aspects of sound systems (Scobbie, 2006). The speaker-specific implementation of VOT targets as shown in the results of the present study seems to offer a challenge to feature-based modular approaches. It was observed that the VOT values are not either distinctive or redundant, but gradient. Each part of the system might be linked more or less directly to the other relevant cues that might pertain.

The present paper is a first step towards a more comprehensive examination of the within-speaker variation. A further systematic study based on a large amount of data would be necessary for a more complete picture of speaker-specific variation.

References

Abramson, A. S. 1972. "Tonal experiments with whispered Thai," In A. Valdkman (Ed.), *Papers on Linguistics and Phonetics to the Memory of Pierre Delattre*. The Hague: Mouton, 29-55.

Bybee, J. 2001a *Phonology and Language Use*. Cambridge: CUP.

Bybee, J. 2001b. "The phonology of the lexicon: Evidence from the lexical diffusion," In S. Kemmer & M. Barlow (Eds.), *Usage Based Models of Language*. CSLI, Stanford University, 65-85.

Cho, Y. Y. 1987. "The domain of Korean sandhi rules," paper presented at the 62nd LSA meeting.

Cho, Y. Y. 1990. "Syntax and phrasing in Korean," *The Phonology-Syntax Connection,* (Eds.), S. Inkelas & D. Zec, Chicago: University of Chicago Press.

Chomsky, N. & Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Flemming, E. 1997. "Phonetic detail in Phonology: Towards a unified account of assimilation and coarticulation," In K. Suzuki & D. Elzinga (Eds.), *Proceedings of the 1995 Southwestern Workshop in Optimality Theory (SWOT),* University of Arizona.

Foulkes, P. & Docherty, G. 2006. "The social life of phonetics and phonology," *Journal of Phonetics* 34, 409-438.

Goldinger, S. 1997. "Words and voices: Perception and production in an episodic lexicon," In K. Johnson & J. Mullennix (Eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, 33-66.

Han, J.-I. 2000. "Intervocalic stop voicing revisited," *Speech Sciences* 7(1), 203-216.

Harrington, J. 2006. "An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts," *Journal of Phonetics* 34, 439-457.

Hawkins, S. 2003. "Roles and representations of systematic fine phonetic detail in speech understanding," *Journal of Phonetics* 31, 373-405.

Hawkins, S. & Smith, R. 2001. "Polysp: A polysystemic, phonetically-rich approach to speech understanding," *Revista di Linguistica* 13, 99-188.

Johnson, K. 1997. "Speech perception without speaker normalization," In K. Johnson & J. Mullennix (Eds.), *Talker Variability in Speech Processing*, San Diego: Academic Press, 145-166.

Johnson, K. 2007. "Decisions and mechanisms in exemplar-based phonology." In M.-J. Sole, P. Beddor, & M. Ohala (Eds.), *Experimental Approaches to Phonology*. Oxford University Press, 25-40.

Jun, S.-A. 1995. "Asymmetrical prosodic effects on the laryngeal gesture in Korean," *Phonology and Phonetic Evidence, Papers in Laboratory Phonology,* IV.

Jun, S.-A. 1996. "Influence of microprosody on macroprosody: a case of phrase initial strengthening," *UCLA Working Papers in Phonetics,* 92.

Kang, O. 1992. *Korean Prosodic Phonology*, Ph.D. dissertation, University of Washington.

Keating, P. 1985. "Universal phonetics and the organization of grammars," In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, Orlando: Academic Press,

115-132.
Kim, M.-R., Beddor, P. S. & Horrocks, J. 2002. “The contribution of consonantal and vocalic information to the perception of Korean initial stops,” *Journal of Phonetics* 30, 77-100.
Kim-Renaud, Y.-K. 1974. *Korean Consonantal Phonology*, Ph.D. dissertation, University of Hawaii.
Labov, W. 1994. *Principles of Linguistic Change. Vol. 1: Internal Factors*, Oxford: Blackwell.
Lee, H.-B. 1974. “Rhythm and intonation in Seoul Korean,” *Language Research* 10:2, Language Research Institute, Seoul National University (written in Korean).
Martens, C. & Martens, P. 1965. *Phonetik der Deutschen Sprache*, Munchen: Max Hueber Verlag.
Mullenix, J. 1997. “On the nature of perceptual adjustments to voice,” In K. Johnson & J. Mullennix (Eds.), *Talker Variability in Speech Processing*, San Diego: Academic Press, 67-84.
Pierrehumbert, J. 2001. “Stochastic phonology,” *Glot International* 5, 195-207.
Pierrehumbert, J. 2002. “Word-specific phonetics,” In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology* VII. Berlin: Mouton de Gruyter, 101-140.
Pierrehumbert, J. 2003 “Probabilistic phonology: Discrimination and robustness,” In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press, 177-228.
Pierrehumbert, J., Beckman, M. & Ladd, R. 2000. “Conceptual foundations in phonology as a laboratory science,” In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological Knowledge: Conceptual and Empirical Issues*. New York: Oxford University Press, 273-304.
Pisoni, D. 1997. “Some thoughts on normalization in speech perception,” In K. Johnson & J. Mullennix (Eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, 9-32.
Scobbie, J. 2006. “Flexibility in the face of incompatible English VOT systems,” *Laboratory Phonology* 8. 367-392.
Shin, J. 2000. *Malsoliuy Ihay* [Understanding Speech Sounds], Seoul, Korea: Hankwukmwunhwasa.
Silva, D. 1992. *The Phonetics and Phonology of Stop Lenition in Korean*, Ph.D. dissertation, Cornell University.
Silva, D. 1994. “The effects of prosodic structure and consonant phonation on vowel pitch in Korean: an examination of bilabial stops,” ms, University of Texas, Arlington.
Silva, D. 2006. “Acoustic evidence for the emergence of tonal contrast in contemporary Korean,” *Phonology* 23, 287-300.
Smith, C. 1997. “Fricative devoicing: effects of prosodic context on a lenition process,” *NELS* 27, 397-411.
Thomas, E. R. 2002. “Instrumental phonetics,” In J. K. Camber, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change*, Oxford: Blackwell Publishers Ltd, 168-200.

received: 2008. 10. 31
revised: 2008. 11. 26
decision: 2008. 12. 10

▲ Jeong-Im Han
Department of English Language and Literature Konkuk University
Hwayang-dong 1, Gwangjin-gu Seoul, 143-701, Korea
Tel: +82-2-450-3339 (O) Fax: +82-2-450-3335
E-mail: jhan@konkuk.ac.kr

▲ Jooyeon Kim
Department of Korean Language and Literature Konkuk University
Hwayang-dong 1, Gwangjin-gu Seoul, 143-701, Korea
E-mail: kjooyeon@konkuk.ac.kr