1999년도 한국음향학회 학술발표대회 논문집 제18권 제2(s)호

한국어의 유성음화 고찰

한 정임
서울대학교

Stop Voicing in Korean: Allophonic or Purely Phonetic?

Jeong-Im Han
Seoul National University
jihan@plaza.snu.ac.kr

1. Introduction

It is widely assumed that voiceless plain stops are voiced between two voiced segments in Korean. This change, which I call the Plain Stop Voicing (PSV) here, is argued to be categorical modification of the target consonants in most earlier analyses based on phonological approach (Cho 1987, 1990; Kim-Renaud 1974; Kang 1992). On the other hand, work based on phonetic experiments address 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 gradience of the lenis stop voicing process, Jun (1995) argues that "Lenis [plain] Stop Voicing in Korean is not a phonological rule. Rather ... a gradient process." (p. 248). The PSV, therefore, can be accounted for from a different angle in these approaches. More specifically, it does not involve the categorical modification of segments at all, and rather better explained with a gestural overlap and reduction (Browman and Goldstein 1986, 1989). However, note that even in these analyses arguing for a phonetic account of the PSV, they don't have any disagreement on that in word-internal position, plain stops are almost completely voiced, for example, 91% in Silva (1992).

This study shows 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 allophones. The results of the experiment reported here thus provide further support for the phonetic interpretation of the Korean voiced stops. To show this, I investigate the word-internal, intervocalic plain stops. The experiments conducted in this study are divided in two parts. In the first part, this 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). These two dimensions are investigated, examining whether there is any speaker variation in values of these two dimensions. In the rest part of this study, the fundamental frequency pattern at the vowel onset after the target consonant is examined to see the status of the PSV. Following Kingston and Diehl (1994), the natives' perception for the phonological category is shown to be reflected in the F0 pattern of vowels following the target consonants. Thus if the PSV, in spite of many characteristics of a gradient process, shows consistency with the fundamental frequency pattern of the vowel onset, we can say that this process is allophonic, namely, phonological.

2. Methods

2.1 Stimuli

The test words consist of VCV sequences were constructed, in which the V segments are /i/, /e/, /ɨ/, /u/, /o/, /ɔ/, or /a/, and C segments were all bilabial stops, as shown in (1). The reason for the place of articulation for stops limited to bilabial position is that this position was more clearly displayed and thus more easily discernable, as compared to other place of articulations (Silva 1992). As for the vowels, /ɛ/ was not included in the test words, following Lee (1974), and Hong (1987), where /e/ and /ɛ/ are merged at the phonetic level among the speakers of younger generation. Each test word was recorded in the frame sentences as in (2).

(1) test words

$V_1/p/V_2$

V_1 = i, e, ɨ, u, o, ɔ, a

V_2 = i, e, ɨ, u, o, ɔ, a

(2) frame sentences

ikɔ irimi _____ ta.

'The name of this is ____'

2.2 Subjects and Recording

Four male speakers, who were students or professors at Seoul National University, 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 in the Language Research Institute at Seoul National University, using a microphone (AKZ, model C414B-ULS) and a high-quality cassette recorder (Tascam, model 122MKⅢ). 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 three times in random order, at a natural, comfortable speed. The test sentences were presented in lists in Korean orthography.

2.3 Analysis

The recorded data were digitized onto a CSL (Computerized Speech Lab) (model 4300) at a sampling rate of 16000 sample/second, and then stored as files to be processed by the commercial software package MULTI-SPEECH (model 3700). Waveforms, spectrograms and pitch tracks for each token were created. The total number of tokens was 587 (49 words x 4 speakers x 3 repetitions). One token from a speaker was not considered in this experiment for its bad resolution in the waveform.

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 spectrogam. The beginning of stop release was captured with a sudden increase in energy in both the 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. As for the measurement of pitch patterns, voicing period marks were placed at the beginning of each vowel period, from the vowel onset to the fifth vowel period. The cues for the beginning of the vowel onset were the beginning of the second formant structure. Then the fundamental frequency values were measured at each point.

3. Results and Discussion

3.1. Experiment one: the likelihood and the amount of voicing

Is there any significant difference in voicing among speakers? It can be examined as to the "likelihood of voicing" and the "amount of voicing". Likelihood refers to how often speakers pronounce the target consonants as fully voiced, which is equal to the number of full voiced tokens out of the whole tokens. On the other hand, the amount of voicing refers to the percentage of voicing portion during the closure duration in each token in each speaker.

First, the results of the experiments for the likelihood of voicing were presented in Table 1.

(Table 1 here)

For each utterance containing a target stop, likelihood was calculated for each speaker by counting the number of repetitions of that utterance that were fully voiced during the whole interval of closure. Table 1 shows that only one speaker shows more than 90% of full voicing as proposed by previous phonetic work as well as phonological analyses. Speaker 1 shows only small number of tokens which are fully voiced. However, more importantly, intervocalic plain stops are realized with a varying degree of voicing depending on speakers in this prosodic position.

Now turn to the issue of the amount of voicing. The assumption that there is no difference in voicing among speakers is tested by comparing the mean values of voicing across speakers. The mean values, standard deviation and standard errors of voicing are reported in Table 2 and Table 3.

(Table 2 and 3 here)

The means are compared pair-wise using the t-test, and the t-values are reported in Table 3. The results indicate that the above assumption is rejected at 1% significance level in all pair-wise comparisons except for that between speakers 2 and 3. It is clearly shown that speakers show quite amount of voicing during the whole stop closure duration, more than 65 % in each speaker. But again the speaker varied considerably as to the amount of voicing during stop closure. This is graphically represented in the following figure.

(Figure 1 here)

In figure 1, although all speakers produced both voiced and devoiced tokens, they varied as to how many of their tokens fell into each of the four voicing categories. Speaker 1 produced the most devoiced tokens, while speaker 2 produced most of tokens as fully voiced.

In sum, the likelihood and the amount of voicing pattern show that even in word-internal, intervocalic position, speakers do not show any consistent pattern in either of dimensions. Speakers vary considerably as to the number of tokens that were fully voiced, and also as to the amount of full voicing during the stop closure. The result of the experiment reported here suggest that the PSV is not a categorical change as assumed in phonological analyses.

3.2. Experiment three: f0 change at the following vowel onset

To investigate further whether the PSV is a phonological process or purely phonetic one, the onset value of fundamental frequency at the following vowel is examined. According to Silva (1994), change in f0 values during the first 10% of the following vowel is significantly different among phonation types: the largest change in f0 occurs in the tense stops which is followed by aspirated and plain stops in this order. Following the phonological analyses, the f0 values show a consistent pattern regardless of the amount of voicing of the preceding consonants, whether the f0 values are slightly rising (Kim and Duanmu 1998) or slightly falling (Silva 1994).

To investigate the change in f0 values (the change in values from f01 to f05) with voicing, it is necessary to define the following 5 types of changes. Here "f01" through "f05" represent the fundamental frequency values of the first five pitch periods of the vowel after the target consonant.

(3) Type 1 Change = f02-f01
Type 2 Change = f03-f02
Type 3 Change = f04-f03
Type 4 Change = f05-f04
Type 0 Change = f05-f01
(= Sum of Type 1 through Type 4)

As shown in Table 4, the f0 values generally decline as they change from f01 to f05, and it is notable that the Type 1 change is most significant. As a result, the Type 1 change accounts for 69% of the total change denoted as Type 0 change.

(Table 4 here)

When these changes are linked to the level of voicing, the f0s are found to decrease less when voicing has a high value. To see the result rigorously, the various types of changes are estimated as a linear function of the voicing levels using the ordinary least squares method. To put differently, I run the following regression for each of the 5 types of the change in f0 values.

(4) Change = $\alpha + \beta *$voicing $+ \varepsilon$, where $\varepsilon \sim N(0, \sigma)$

The results are reported in Table 5.

(Table 5 here)

The results in Table 5 indicate that most types of changes are positively associated with the level of voicing: both the correlation coefficients and most of β estimates are positive. Given than all types of the changes are negative (See Table 4), the positive association implies that a higher value of voicing reduces the decline -- that is, the f0s decline by less when voicing takes a high value. It is notable that such positive association between the change and the level of voicing is significant only for the Type 1 change, the change from f01 to f02. The regression results indicate that, if the speaker variation is uncontrolled for, a one unit increase in voicing reduces the decline from f01 to f02 by 0.246 units. Given that a similar increase in voicing reduces the decline from f01 to f05 (the overall change) by 0.270 units, the change at the first stage accounts for 91.1% of the total changes. In contrast, the changes of other types are almost unrelated to the level of voicing. The results do not qualitatively change when the speaker variation is controlled for, when the regression coefficients are somewhat smaller.

It is also notable that the changing patterns of f0 values vary among speakers. The patterns are described in Figure 2, which plots the value of f01-f05 by speakers. The horizontal axis of the diagram represents speakers, each denoted by a number from 1 through 4. The vertical axis of the diagram measures the mean values of f0 values for each speaker, where each of f01-f05 is differently shaded.

(Figure 2 here)

The results of the experiment for the f0 change show that the f0 pattern at the initial position of the following vowels is not consistent in terms of speakers. Again this can be a further support for the non-categorical characteristic of the PSV.

Overall, even in word-internal, intervocalic position, the likelihood and the amount of voicing during closure are not consistent enough; the amount of voicing is greatly affected by adjacent vowels; the f0 values correspond to the amount of voicing quantitatively, without any consistent pattern. Thus the PSV appears to be purely phonetic process, not allophonic as assumed in phonological analyses.

Thus the characteristic of gradient voicing supports the interpretiona that the PSV is not a phonological change.

4. Conclusion

From the results of the acoustic experiments presented here, it would be possible to conclude that the PSV is a purely phonetic process, not allophonic as assumed in phonological analyses. Even in word-internal, intervocalic position, both the likelihood and the amount of voicing during stop closure are not consistent enough among speakers; the voicing pattern is greatly influenced by adjacent vowels; the f0 values correspond to the amount of voicing quantitatively, without any consistent pattern.

References

Browman C. and L. Goldstein (1986) "Towards an articulatory phonology," *Phonology 3*.

Browman C. and L. Goldstein (1989) "Articulatory gestures as phonological units," *Phonology 6*.

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. by) S. Inkelas and D. Zec, Chicago: University of Chicago Press.

Hayes, B. and T. Stivers (1996) "A phonetic account of postnasal voicing," ms. UCLA.

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 strenthening, *UCLA Working Papers in Phonetics 92*.

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

Kim, M.-R. and S. Duanmu (1998) "The correlation between onset types and tone in Chonnam Dialect," *Proceedings of the ICKL 11*.

Kim-Renaud, Y.-K. (1974) *Korean consonantal phonology*, Ph.D. dissertation, University of Hawaii.

Kingston, J. and R. Diehl (1994) "Phonetic knowledge," *Language 70*.

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.

Smith, C. (1997) "Fricative devoicing: effects of prosodic context on a lenition process," NELS 27, 397-411

Table 1. The overall percentage of tokens produced by each speaker that were fully voiced

speaker	percentage of voicing	total number of tokens	tokens which were voiced
speaker 1	34%	146	49
speaker 2	74%	147	109
speaker 3	78%	147	115
speaker 4	95%	147	139

Table 2. The mean phonetic values and standard errors of voicing during stop closure among various speakers (ms).

	Speaker 1	Speaker 2	Speaker 3	Speaker 4
Mean	66.93	87.43	90.79	93.31
(Standard Deviation)	(27.26)	(22.89)	(19.06)	(7.41)
(Standard Error)	(2.26)	(1.89)	(1.64)	(0.61)

Table 3. The statistical significance of the mean voicing values at each preceding vowel.

	Speaker 2	Speaker 3	Speaker 4
Speaker 1	-8.55 (0.00)	-9.97 (0.00)	-13.09 (0.00)
Speaker 2		-1.41 (0.16)	-4.55 (0.00)
Speaker 3			-3.14 (0.00)

Note: i) The t-values are reported, and those in parentheses are the p-values.
ii) The t-values are based on the mean value of the speaker in the column minus that of speaker in the row.

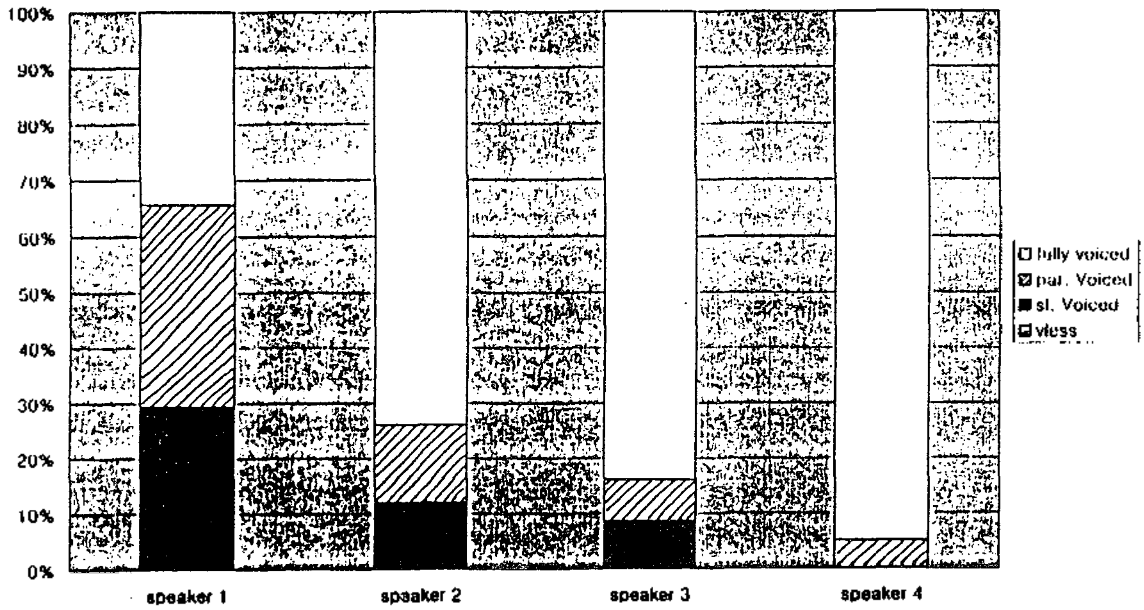

Figure 1. The overall percentage of tokens produced by each speaker that were voiceless, slightly voiced, partially voiced, and fully voiced.
(fully voiced = 100% of voicing, partially voiced = 51-99% of voicing, slightly voiced = 1-50% of voicing, voiceless = 0% of voicing)

Table 4. The mean values of changes of various types (Hz).

	Type 0	Type 1	Type 2	Type 3	Type 4
mean (s.d.)	-14.37 (19.64)	-9.87 (18.64)	-2.40 (8.27)	-0.78 (8.95)	-1.32 (8.97)

Table 5. The relationships among the change in f0s and voicing.

		Type 0	Tyep 1	Type 2	Type 3	Type 4
correlation coefficient with voicing		0.323	0.310	0.030	0.004	0.031
Regression	No Speaker Dummy[1)]	0.270[3)] (0.033)	0.246[3)] (0.031)	0.010 (0.015)	0.002 (0.016)	0.012 (0.016)
	Speaker Dummy[2)]	0.193[3)] (0.034)	0.182[3)] (0.033)	-0.001 (0.016)	0.015 (0.018)	-0.003 (0.018)

Note: i) The potential differences among speakers in f0s and voicing are ignored, and the estimated equation is: each change= $\alpha + \beta *\text{voicing} + \varepsilon$.

ii) Any potential differences among speakers in f0s and voicing are controlled for with speaker-specific effect (η_s), and the equation is: each change= $\alpha + \beta *\text{voicing} + \eta_s + \varepsilon$, where s denotes speaker.

iii) Differs from 0 with 1% statistical significance.

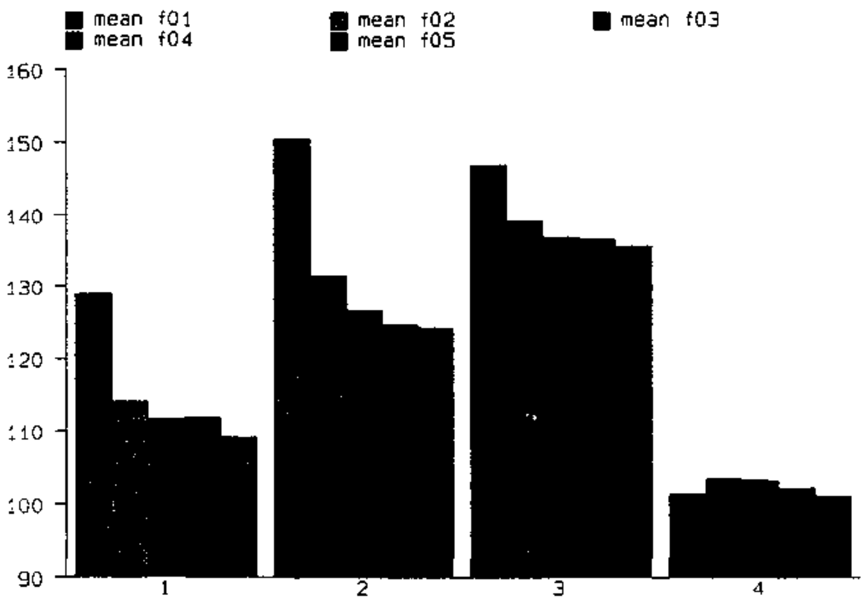

Figure 2. The mean values of f01-f05 by speakers