PRODUCTION AND PERCEPTION OF WORD PROSODY IN THREE DIALECTS OF KOREAN

Kenji Yoshida, Junghyoe Yoon, and Hyun-jin Kim

Indiana University, Bloomington
keyoshid@indiana.edu, junyoon@indiana.edu, hk14@indiana.edu

ABSTRACT

This paper examines the relationship between production and perception of prosodically marked lexical contrast, comparing 16 native speakers from three dialects of Korean known to exhibit variation in the use of prosodic features for lexical marking. A set of synthesized stimuli was constructed, where both F0 contour and syllable duration were manipulated. South Kyungsang speakers have F0 distinction in production and are sensitive to F0 variations in perception. Cholla speakers are sensitive to F0 in the opposite direction to Kyungsang speakers, suggesting that their 'interpretation' of the F0 is the critical factor of perceptual judgment. Some of the Seoul speakers show a duration contrast and are sensitive only to duration change. The results reveal general though incomplete correlation between production and perception of word prosody, suggestive of the different typological status of the three dialects.

Keywords: Word prosody, length contrast, pitch accent, dialects, Korean.

1. INTRODUCTION

Models of speech assume certain 'links' between production and perception of speech sounds. Some appear to conceive that the two are largely the same, while others claim that the two are independent from each other, but show correlation via interaction of speakers and listeners [5]. This question becomes especially important when production and perception are expected to show some discrepancy. In this light, dialects sharing similar but slightly different sound systems are one potential case. The present study examines the production and perception of pairs of words in Korean reported to exhibit prosodic variation depending on dialects.

While Korean historically has quantity contrasts, there is evidence that the standard, Seoul Korean is losing it [7]. But this distinction is apparently retained in the southern parts of the country, such as Cholla-do [3] and Kyungsang-do [2]. Seoul Korean has an intonational system with phrasal tones but no lexical pitch accent [2, 3], while most Kyungsang-do dialects have lexical accent systems [2]. The present research explores how speakers of three dialects mark the lexical contrast and how their different production patterns correlate with their perception of the contrast.

2. PROCEDURE

This study consists of two parts; a production experiment and a perception experiment. In both, words constituting minimal pairs in most dialects were used as in Table 1. In South Kyungsang dialects (SK), they expected to show both duration and pitch contrast, whereas in Cholla dialects (CL), only a length contrast is expected and for Seoul dialects, no contrast is expected [2, 3].

Table 1: Lexical Corpus

Meaning	Seoul	Cholla	Kyungsang
snow	nun	nu:n	nu:n (low)
eye	nun	nun	nun (high)
chestnut	pam	pa:m	pa:m (low)
night	pam	pam	pam (high)

2.1. Participants

16 speakers of three dialects participated in both production and perception experiments, 5 speakers from Seoul, 6 from CL (all from Gwanju), and 5 from SK (3 from Busan and 1 from Jinju and Changwon). The participants of Seoul and CL dialects were recruited in Seoul. CL speakers have spent 2 to 5 years in Seoul. SK speakers were recruited in their own cities. Two of the SK subjects were in their late 50's; others were in their late 20's to early 30's. The participants were paid for participation. No self-reported hearing impairment was reported.

2.2. Production experiment

The test words were embedded in frame sentences each containing both contrasting words (1a-2b).

Taking advantage of *scrambling*, the location of the target words were alternated, yielding two pairs of sentences with largely the same meaning.

(1a) [tɕɔlsu-gɑ **nun**-ɯl **nun**-e nɔ-ɔt-ta]
Chulsoo-nom. snow-acc. eye-loc. put-past-decl.
(1b) [tɕɔlsu-gɑ **nun**-e **nun**-ɯl nɔ-ɔt-ta]
Chulsoo-nom. eye-loc. snow-acc. put-past-decl.
' Chulsoo put snow onto the eyes.'
(2a) [tɕɔlsu-gɑ **pam**-ɯl **pam**-e mɔg-ɔt-ta]
Chulsoo-nom. chestnut-acc. night-temp. eat-past-decl.
(2b) [tɕɔlsu-gɑ **pam**-e **pam**-ɯl mɔg-ɔt-ta]
Chulsoo-nom. night-temp. chestnut-acc. eat-past-decl.
' Chulsoo ate chestnuts at night.'

The sentences were read at least five times in non-randomized order to avoid confusion between two meanings, yielding 10 tokens for each target word (5 in initial and second position, respectively). The recordings were made in quiet rooms, such as university labs using a head-mounted microphone (SONY ECM-140). Their speech was directly recorded onto a laptop PC and digitized using Audacity at 16 KHz with 16 bit quantization.

2.3. Perception experiment

Based on a preliminary experiment with 2 speakers from SK and CL dialects, which provided virtually the same results as in this paper, perception stimuli were constructed by manipulating two aspects of the signal, F0 contours and duration. The stimuli then were embedded in the sentence (3), with a stop in the middle of the utterance.

(3) [tɕɔlsu-gɑ nun-ɯl]
Chulsoo-nom. 'nun'-acc. …

The task of the participants was to guess which of the alternatives (4a, 4b) given on a test sheet is likely to follow based on what they heard. It indicates the choice between two possible meanings of the syllable /nun/ ('snow' or 'eye').

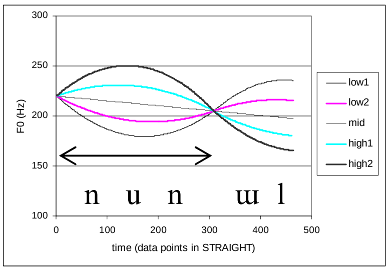
(4a) [mɑdɑŋ kusɔg-ɯro tɕiw-ɔt-tɑ]
yard corner-to remove-past-decl.
'Chulsoo removed (snow) to the corner of a yard.'
(4b) [mul-lo s'is-ɔt-ta]
water-with wash-past-decl.
'Chulsoo washed (his eyes) with water.'

An original speech token of a speaker from SK dialect (27 yrs. female, Jinju-city) was chosen and the target word /nun/ excised. The sound clip was analyzed and its F0 contour and duration were manipulated using STRAIGHT [3]. The contour was made to vary in 5 steps from the one with the lowest F0 excursion (179.6 Hz), through a flat line (with a declination slope), and finally to the highest F0 excursion (250.4 Hz) in equal intervals of approximately 1.0 semitone. F0 contours were computed with parabolic approximation of the observed F0 contours, given by the formula (5).

$$(5)\ f(t) = a \cdot (t - h)^2 + k + s \cdot (t - 1)$$

t is the number of data points in time. *a* is a constant of the magnitude of F0 excursion determined by the desired F0 extreme values. *h* is a constant for the mid point of the target syllable. *k* is the F0 value at the extreme point. *s* is another constant of declining slope interpolating the starting point and ending point of the syllable. The F0 contour of the following particle /ɯl/ was also manipulated to get smooth and natural F0 continuation. The resulting F0 contours are shown in Figure 1. In addition to this, the duration of the target word (indicated by arrow in Figure 1) was proportionally shortened and lengthened in 4 steps, short, normal, long, extra long (0.65, 1.00, 1.35 and 1.70 times the original duration). The 20 stimuli thus created were sorted in 10 randomized orders and presented to the listeners through a headset. In total, 200 responses per participant were obtained.

Figure 1: Five F0 contour of synthesized stimuli.



3. RESULTS

3.1. Experiment I -- Production

To examine how speakers differentiate the minimal pairs, F0 and vowel duration of the target words were measured using Praat. The F0 value of the target vowel 40 ms. prior to the end of the nucleus vowel was measured. This time location allows us to avoid the difficulty of determining the mid point of the vowel when the vowel onset is blurred by strong aspiration of lax [p], especially for Seoul and CL speakers. The peak F0 was not chosen too, as some target words have concave pitch contours. Aspiration was included in vowel duration.

Table 1: Result of 2-way ANOVA (meaning, location, meaning*location), showing only **main effect** of 'meaning' on **vowel duration**. Shaded cells indicate that the main effect was not significant (α = .05).

Seoul	/pam/			/nun/		
Subjects	df	F	p	df	F	p
S1	1, 16	9.95	.006	1, 16	9.80	.006
S2	1 ,15	20.95	.000	1, 16	0.43	.524
S3	1, 16	0.04	.839	1, 22	12.33	.002
S4	1, 19	0.64	.434	1, 14	0.05	.821
S5	1, 20	0.38	.544	1, 19	3.62	.072
Cholla	/pam/			/nun/		
Subjects	df	F	p	df	F	p
C1	1, 16	1.08	.314	1, 16	4.11	.060
C2	1 ,16	2.46	.137	1, 16	0.10	.755
C3	1, 14	13.49	.003	1, 16	67.60	.000
C4	1, 16	0.39	.543	1, 16	0.06	.804
C5	1, 16	3.47	.081	1, 14	17.99	.001
C6	1, 24	267.72	.000	1, 16	187.62	.000
Kyungsang	/pam/			/nun/		
Subjects	df	F	p	df	F	p
K1	1, 39	140.56	.000	1, 36	61.71	.000
K2	1 ,16	7.03	.017	1, 16	2.91	.108
K3	1, 15	.06	.805	1, 16	3.96	.064
K4	1, 16	1.05	.321	1, 16	.046	.832
K5	1, 30	1.48	.233	1, 30	.564	.458

Table 2: Result of 2-way ANOVA, showing only **main effect** of 'meaning' on **vowel mid F0**.

Seoul	/pam/			/nun/		
Subjects	df	F	p	df	F	p
S1	1, 15	0.27	.609	1, 16	5.41	.033
S2	1 ,16	3.14	.097	1, 16	1.13	.285
S3	1, 16	1.18	.294	1, 22	2.03	.168
S4	1, 19	0.01	.943	1, 14	0.00	.992
S5	1, 20	0.52	.479	1, 19	0.64	.432
Cholla	/pam/			/nun/		
Subjects	df	F	p	df	F	p
C1	1, 16	96.34	.000	1, 16	124.16	.000
C2	1 ,16	1.17	.296	1, 16	1.19	.292
C3	1, 14	0.41	.530	1, 16	62.83	.000
C4	1, 16	0.02	.905	1, 16	0.90	.357
C5	1, 16	2.40	.141	1, 16	4.36	.053
C6	1, 24	0.23	.636	1, 16	35.03	.000
Kyungsang	/pam/			/nun/		
Subjects	df	F	p	df	F	p
K1	1, 39	344.07	.000	1, 36	274.16	.000
K2	1 ,16	29.78	.000	1, 16	4.92	.041
K3	1, 15	197.65	.000	1, 16	172.19	.000
K4	1, 16	18.16	.001	1, 16	14.27	.002
K5	1, 30	8.38	.007	1, 30	5.33	.028

Measurements were analyzed separately for each subject. 2-way ANOVAs were conducted with ‘meaning’ and ‘sentence position’ as factors. Table 1 and 2 summarize the main effects of the lexical meaning of target word, with unshaded boxes indicating significant results. The most consistent result is that all SK speakers had a significant F0 contrast, while only a few speakers in other dialects show F0 contrast. Unlike our expectation, only half of the CL speakers have a significant contrast in duration, and two Seoul speakers show significant durational contrasts.

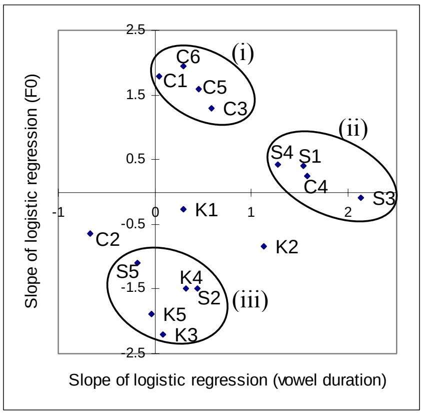
3.2. Experiment 2 --Perception

The lexical judgments of the listeners ('snow' or 'eye') are grouped according to the type of stimuli they heard. We report results for F0 and duration, collapsing across levels of the other variable. As the dependent variable is binary and the discrimination function is expected to be non-linear, the sharpness of the discrimination function is computed with logistic regression [8], given as the coefficient term b in (6).

(6) $\ln(\frac{P}{1-P}) = a + bx$

Figure 2 plots the result for duration and F0. Three separate groups are identified as indicated by ellipses (i-iii), which largely coincide with the listeners’ dialects. Listeners (iii) are selectively sensitive to F0 information, as expected for SK listeners (low = 'snow', high = 'eye'). Listeners (i), all consisting of CL listeners, are sensitive to F0, but in the opposite direction to (iii), although they tend to show both F0 and duration contrast in production. Listeners (ii) (mostly Seoul) are sensitive to duration information (long = 'snow').

Figure 2: Slope of discriminating functions (coefficient b of logistic regression) for vowel duration (abscissa) and F0 (ordinate).

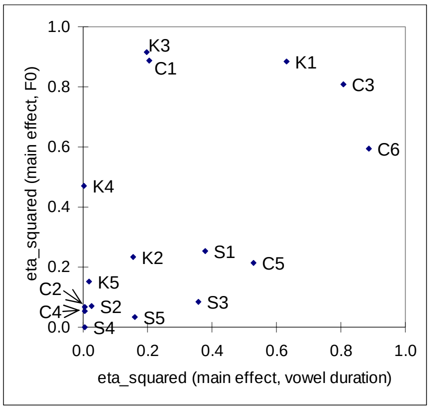


4. DISCUSSION AND CONCLUSIONS

4.1. Production and perception

To examine the correlation between production and perception, we take up a measure of how sharply the participants differentiate the word pairs in production, an effect size estimate η^2 (eta-squared, proportion of variance explained) for the ANOVA factor 'word meaning'. Note that η^2 is a standardized measure, varying from 0 to 1. Large η^2 value indicates a large effect size, thus clearer distinction. η^2 is plotted in Figure 3 for duration and F0.

Figure 3: Effect size (η^2) of main effect by 'lexical meaning of the target words' on vowel duration (abscissa) on F0 (ordinate) for /nun/ pairs.



Both group (i) and (iii) in Figure 2 are those with large η^2 in the F0 analysis. Although they are not separated in η^2 space, the actual F0 patterns are totally different. SK has clear contrast of low ('snow') and high ('eye') F0 extremes, while CL speakers have rising F0 contour for both words. Further acoustic analysis, however, reveals that one of the two ('eye') has later peak F0 relative to the nuclear vowel (for C1, C3 and C6, p=.000), which is exactly the one associated to the stimuli with lower F0 in lexical judgments. This explains the opposite direction of the perceptual discrimination function from SK. While SK participants in (iii) responded to F0 information associated to lexical contrast, CL speakers in (i) interpret concave F0 contours as an indication of late rise and gave 'eye' judgments. Two of three Seoul listeners in (ii) showed a contrast in duration and responded to durational information. Some participants were not in these groups, perhaps for reasons specific to individuals (e.g, K2 seems to be in the stage of migrating from (iii) to (ii), K1 appeared to fail to follow the experimental task).

The results reveal a certain degree of correlation between production and perception. Of the two aspects, perception appears to be more revealing of distinct traits. Some speakers did not reveal lexical contrast in production possibly because they have learned to suppress their dialectal traits in production. If this speculation is correct, it is remarkable that 4 out of 6 CL speakers show their distinctive characteristics in perception.

4.2. Implication for typology of word prosody

The present results indicate a large disparity in word prosody between the dialects that share basic lexical items and grammar. The SK dialect has tonal and accentual characteristics, showing lexical contrast in F0 in both production and perception. The CL dialect has quantity-based prosody, showing lexical contrast in duration. However it also shows a sign of 'early vs. late peak' contrast. This may have emerged from difference in the alignment of phrasal tone to segments due to durational difference, which may eventually be reanalyzed as real tonal distinction, in the similar manner to the emergent stress contrast for Seoul Korean [1]. Lastly, despite the emergent trend in other phonological classes [3, 6], for the monosyllable words examined here, Seoul dialect appears to maintain its non-tonal and non-accentual characteristics, showing no tonal contrast in production or perception. Instead, some of them exhibit durational distinction, suggestive of quantity-based word prosody.

5. REFERENCES

[1] de Jong, K. 2000. Attention modulation and the formal properties of stress systems, *CLS* 36, 71-91.

[2] Fukui, R. 2001. Pitch accent systems in Korean. *J. Phonet. Soc. Jpn.* 5-1, 11-17. (in Japanese)

[3] Jun, S. A. 1996. *The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure*. NY, Garland.

[4] Kawahara, H., Masuda-Katsuse, I., and Cheveigne, A. 1999. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds, *Speech Commun*. 27, 187-207.

[5] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling*, Dordrecht, Kluwer, 403-439.

[6] Silva, D. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23, 287-308.

[7] Umeda, H. 1995. Kankokugo no boin (Vowels of Korean). *J. Ling. Soc. Jpn*. 106, 1-21. (in Japanese)

[8] Xu, Y., Gandour, J. T. and Francis, A. L. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.* 120(2), 1063-1074.