The Perception of Korean Boundary Tones by First and Second Language Speakers

Hae-Sung Jeon

School of Language, Literature and International Studies, University of Central Lancashire, UK

HJeon1@uclan.ac.uk

Abstract

This paper reports an experiment which investigated the perception of prosody in Korean or non-word utterances by native Korean speakers and English learners of Korean. Listeners rated the degrees of positivity and excitement of resynthesized utterances with different pitch ranges and durations. The results revealed no significant differences between the two groups of listeners. The variations in pitch range and duration had systematic effects on the ratings. However, the interactions between various factors suggest that the mapping between prosodic shapes and their paralinguistic meaning is not straightforward.

Index Terms: prosody, second language, emotion

1. Introduction

Intonation has linguistic and paralinguistic (e.g. emotional, attitudinal) functions which are important in social interactions [1]. There seems to be much cross-linguistic similarity in the paralinguistic (more specifically 'emotional') meaning of prosody; for example, a large pitch movement or an overall increase in the acoustic parameters is related to emotional arousal, e.g. [2]. However, cross-dialectal or -language differences in linguistic or paralinguistic uses of intonation are reported [e.g. 3, 4, 5]. Second language (L2) learners need to acquire the phonology of their target language intonation. Also, linguistic or paralinguistic meaning of prosody can be misinterpreted by speakers of foreign languages.

This study was motivated by English learners' impressionistic comments that Korean prosody often sounds like 'child-whining'. This is probably due to the frequent use of complex boundary tones and significant phrase-final lengthening. Jun [6] identifies nine Intonational Phrase (IP) boundary tones in Korean, and previous studies reported significant lengthening in the IP- or utterance-final syllable, i.e. the IP- or utterance-final syllable is longer than the medial syllable by a factor of 1.6-1.8 (see Ch. 2., [7] for a review). The previous studies mainly dealt with declarative sentences ending with a L or H tone; final-lengthening is likely to be even greater when there are multiple tones associated with the right edge of the phrase or utterance.

The aims of the study were to investigate: 1) whether emotional interpretation of prosody differs between native Korean speakers and English speakers learning Korean; 2) the potential effect of the lexical cue on the emotional interpretation of prosody; 3) the relationship between utterance-final lengthening or pitch range expansion and the emotional interpretation of prosody; and 4) whether different boundary tones are associated with different emotional interpretations.

2. Boundary tones in Korean

The nine tones in [6] include L%, H%, LH%, HL%, LHL%, HLH%, LHLH%, HLHL% and LHLHL% which are transcribed at the L(ow) and the H(igh) targets in the F0. These boundary tones are associated with the right edge, particularly the final one or two syllables, in the IP or utterance. Although [6] and [8] list the tone types and their use, they acknowledge that it is difficult to pinpoint the basic meaning of each boundary tone, since the same shape of pitch movement could signal different meanings depending on the context (e.g. [9]).

The descriptions of intonation in [6] and [8] (the transcription in [8] follows the British Framework) can be summarised as follows:

- L%: declaratives; neutral, formal, as-a-matter-of-fact, polite-apologetic
- H%: seeking information as in yes/no questions
- LH%: questions, continuation rises, explanatory endings in declaratives; annoyance, irritation, disbelief
- HL%: declaratives, wh-questions; kind, polite, apologetic
- LHL%: intensifying the meaning of HL%; persuasive, insisting, confirmative; annoyance, irritation
- LHLH%: intensifying some of the meaning of LH%, i.e. annoyance, irritation, disbelief
- HLHL%: intensifying the meaning of confirming and insisting on one's opinion; nagging, persuading
- LHLHL%: similar to LHL%, signalling more annoyance

That is, L% and H% are described as emotionally relatively neutral in a statement or a question respectively, and it is only HL% which may signal speakers' positive feelings in declaratives. Complex tones are associated with irritation or annoyance; similarly, [10] suggests that the tones with three or more targets do not have a distinct meaning of their own but they are used to intensify the meaning of the less complex tones, e.g. HLHL% intensifies the meaning of HL%.

In this experiment, eight boundary tones, L%, H%, LH%, HL%, LHL%, LHLH%, HLHL% and LHLHL%, were employed. These were chosen intentionally to balance the number of relatively simple tones (L%, H%, LH% and HL%) and complex tones (LHL%, LHLH%, HLHL% and LHLHL%). HLH% was excluded in the main experiment, since the stimuli with this tone sounded unnatural compared to the others. HLH% was used in the practice session preceding the main experiment, together with HLHLH% (see Section 3. 1. 2).

3. Experiment

3.1. Method

3.1.1. Materials

There were four types of trisyllabic stimuli, two Martian (non-word) utterances and two Korean utterances. For the Korean materials which were created first, the experimenter (a trained phonetician, female native Korean speaker) recorded two utterances, [migugɛ] ('in the USA' which can be either declarative or interrogative) and [hatsima] ('don't do it', imperative or interrogative) in the ten boundary tones (L%, H%, LH%, HL%, LHL%, LHLH%, HLHL% and LHLHL% for the main experiment; HLH% and HLHLH% for the practice sessions). The recording took place in a sound-attenuated booth in the Phonetics Laboratory, University of Cambridge (digital recording with a sampling rate of 44.1 kHz). Then all resynthesis was performed with [11] using the PSOLA method. The F0 contours of the recorded utterances were stylised to create source utterances for further resynthesis; the L and H targets were identified and the F0 between two adjacent targets was interpolated. The experimental materials were created by resynthesising F0 (PITCH, two levels; Natural, Expanded) and DURATION (three levels; Shortened, Neutral, Lengthened). For the Natural condition, no additional manipulation of F0 was done. To create the Expanded condition, the H peaks were increased by 2 ST; the only exception was with L% in which the valley was lowered by 2 ST. For the Lengthened duration condition, the utterance-final vowel was lengthened by 50%; for the Shortened condition, the vowel was shorted by 50%; the original duration was maintained in the Neutral condition.

The Martian stimuli ([sesapʰa], [samuda]) were created by concatenating syllables from Finnish utterances spoken by a female speaker. None of the syllables were adjacent to each other in the original utterances. After the concatenation, their pitch and duration properties were resynthesised to be identical to the Korean counterparts ([samuda] and [migugɛ], [sesapʰa] and [hatsima]). In Korean, phrase-initial [s, h] are often associated with H and [m] with L [6]. However, Korean listeners are insensitive to the violation in the tone-segment type association [12] and all stimuli sounded natural.

The paper-and-pencil method was used to collect listeners' responses. A two dimensional model of emotion with valence (i.e. the degree of positivity) and arousal (i.e. the degree of excitement) was employed, with answer sheets from the Self-Assessment Manikin (SAM) rating system [13]. This method is particularly useful for cross-cultural studies when the interpretation of emotional words can be culture-specific. Although there is a third dimension, dominance, it was not included since its reliability has been questioned and it was necessary to keep the experiment simple. The answer sheets had pictures representing five steps of valence (the continuum between smiling face and sad face) and five steps of arousal (the continuum along the excitement scale). Listeners could put a mark on a picture or between two pictures, and therefore a 9-step scale for each dimension was provided. High scores for valence and arousal indicate a high degree of happiness and excitement, respectively. Following the descriptions in [6], [8], and [10], it is expected that the complex tones will show lower positivity (valence) and higher excitement (arousal) than the simple tones, although HL% is likely to be positive with lower excitement.

3.1.2. Experimental procedure

Thirty eight participants without speech or hearing problems participated in the experiment (19 native Korean speakers at the Korea National University of Education, South Korea, age Mean = 21.68 years, SD = 2.75; 19 native British English speakers at the University of Central Lancashire, UK, age Mean = 20.68, SD = 1.59). English speakers were students in a Korean language class at beginner level who had been learning Korean for six months for 2 hours a week. Although their command of Korean was at a basic level, they could understand the experimental materials presented in Korean.

All participants in the same first language (L1) group were tested together. Participants sat in a quiet classroom and the stimuli were played on loudspeakers connected to a PC. Participants were told that they would hear a Martian speaking and their task was to guess the degree of positiveness/negativeness (valence) and the degree of excitement (arousal) from the speech. They were instructed to make full use of the nine-degree scale. They were informed that there would be Korean utterances only after the completion of the Martian part. The presentation order was: [samuda], [sesapʰa], [migugɛ] 'in the USA' and [hatsima] 'don't do it'. Within each stimulus type, the main experiment was preceded by a short practice session which allowed participants to familiarise themselves with the task. In the practice sessions, stimuli with HLH% and HLHLH% in two pitch conditions (Neutral, Expanded) with natural duration were played. Each stimulus was played three times and the presentation order was randomised. The experiment took approximately 35 minutes. In total, 192 utterances were played in the main experiment (4 STIMULUS TYPE × 8 TONE TYPE × 3 DURATION × 2 PITCH).

3.2. Results

3.2.1. Overall patterns

Figures 1-4 show the ratings for each TONE TYPE. Overall, there is little difference between the two L1 groups (as supported by the ANOVA results reported below), although the valence ratings seem slightly higher (more positive) in the English group. For valence, the ratings were relatively high (more positive) for the simple tones (L%, H%, LH%, HL%) than for the complex tones. The valence is particularly high for H%, which was probably interpreted as a yes-no question with a relatively neutral emotional state. As predicted, HL% has a high rating for valence; however, contrary to the description in Section 2, LH% also shows high valence. For arousal, the complex tones (LHL%, LHLH%, HLHL%, LHLHL%) were rated higher (more excited). Pitch range expansion (Figs 1 and 2) led to a higher degree of valence and arousal than the natural pitch contour, although the arousal rating seems less affected than valence rating. For duration (Figs 3 and 4), lengthening seems to be associated with lower valence and higher arousal.

A series of mixed ANOVAs were conducted for valence and arousal respectively, with a between-subject factor, L1 (Korean, English), and within-subject factors, STIMULUS TYPE (ST: [migugɛ], [hatsima], [samuda], [sesapʰa]), TONE TYPE (TT: L%, H%, LH%, HL%, LHL%, LHLH%, HLHL%, LHLHL%), DURATION (D: Neutral, Short, Long) and PITCH RANGE (P: Natural, Expanded) (Greenhouse-Geisser correction applied, α level adjusted to 0.01 after Bonferroni correction).

The L1 main effect was not significant as for either valence or arousal. For valence, complicated interactions involving ST were statistically significant together with the main effects of TT, P and D; significant effects were revealed with ST, TT, P, D, ST × TT, TT × P, ST × D, ST × T × D, and P × D. The mean rating was highest for [sesapʰa] (Martian, M = 4.76, SE = 0.14) and lowest for [hatsima] 'don't do it' (Korean, M = 4.1, SE = 0.12) ([migugɛ] M = 4.73, SE = 0.83; [samuda] M = 4.42, SE = 0.12). For arousal, again, the interaction effects involving ST were significant (ST: [migugɛ] M = 4.73, SE = 0.16; [hatsima] M = 4.87, SE = 0.14; [samuda] M = 4.78, SE = 0.14; [sesapʰa] M = 4.68, SE = 0.15). The statistically significant effects included TT, P, D, ST × TT, TT × P, ST × T × P, ST × D, TT × D, ST × TT × D, TT × P × D and ST × T × P × D. Due to the interactions involving ST, further analyses were carried out with the data split by ST ([migugɛ], [hatsima], [sesapʰa], [samuda]). The results of the further ANOVAs are summarised in Table 1. The table shows that there are few effects involving L1, whilst the effects of TT, P, D and TT × D tend to be significant across the STs. Figures 5 and 6 allow an examination of the TT × D interaction.
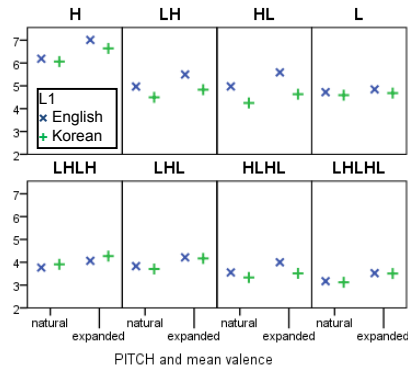

Figure 1: *Valance rating for PITCH. Panels in descending order for mean rating.*
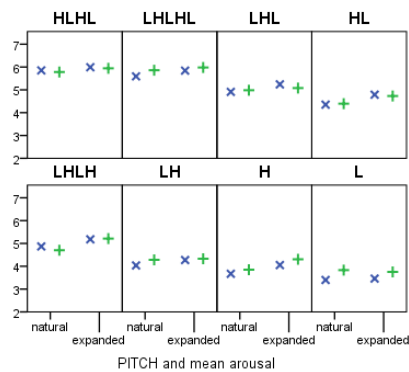

Figure 2: *Arousal rating for PITCH. Panels in descending order for mean rating.*

In Figure 5, [hatsima] ('don't do it') shows little variation across the DURATION conditions with relatively low valence ratings. The other Korean utterance, [migugɛ], shows patterns similar to the Martian utterances in general. Other notable points are: for H%, the valence rating is consistently high for all STs; and for HLHL%, the Korean materials show differences from the Martian materials in that little variation across the DURATION conditions is observed in the Korean materials.
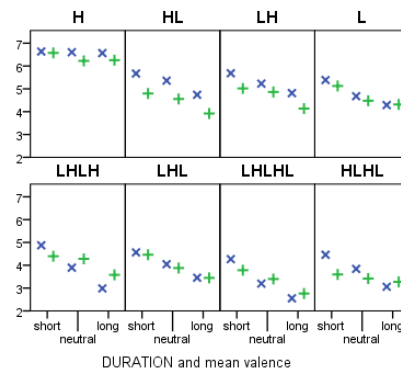

Figure 3: *Valence rating for DURATION. Panels in descending order for mean rating.*
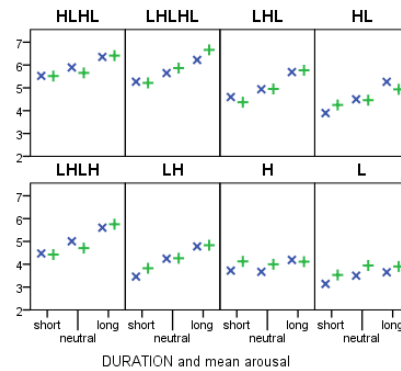

Figure 4: *Arousal rating for DURATION. Panels in descending order for mean rating.*

	valence				arousal			
	K1	K2	M1	M2	K1	K2	M1	M2
TT	**	**	**	**	**	**	**	**
L1	-	-	-	-	-	-	-	-
TT × L1	-	**	-	-	-	-	-	-
P	**	**	**	**	**	-	**	-
P × L1	-	-	-	-	-	-	-	-
D	**	-	**	**	**	-	**	**
D × L1	-	-	-	-	-	-	-	-
TT × P	-	-	-	-	-	-	*	-
TT × P × L1	-	-	-	-	-	-	-	-
TT × D	**	-	*	**	**	*	**	-
TT × D × L1	-	-	-	-	-	-	*	-
P × D	-	-	-	-	-	-	-	-
TT × P × D	-	-	-	-	-	*	**	-
P × D × L1	-	-	-	-	-	*	-	-
T × P × D × L1	-	-	-	-	-	-	-	-

Table 1. *ANOVA results (α level = 0.01), ST: K1 [migugɛ], K2 [hatsima], M1 [samuda], M2 [sesapʰa]. significance codes: ** p < 0.001, * p < 0.01, and - not significant*

In Figure 6, there seems to be a positive correlation between the degree of lengthening and the arousal rating, particularly for Martian, although there are exceptions such as the results for H% and LHL%. For Korean, although the two STs show similar patterns in general, listeners' response patterns differed notably between the two for L%, HL% and HLHL%.

4. Discussion

4.1. Effect of L1

The results show that the emotional interpretation of prosody did not differ between native Korean speakers and native English speakers. English speakers were leaners of Korean at beginner level with a very limited command of Korean and they had been attending the classes for 6 months.

Although the English speakers' valence ratings appeared generally higher than those of the Koreans, the between-L1-group difference did not reach statistical significance. The results seem to support the theory that there is a cross-culturally shared component in the paralinguistic interpretation of prosody related to the expansion of pitch range and durational variations (e.g. [2]). But the generally higher valence ratings by English speakers and the slightly different response patterns between the two groups in some cases (e.g. HLHL% in Fig 3) show that there may also be a language-specific part to be learned which deserves further investigation. Another possibility is that English learners were quick at acquiring the paralinguistic meaning of the Korean boundary tones. Although prosody at the sentence or discourse level is little discussed in the language classroom, students had been exposed to authentic Korean speech through TV shows, films, etc.

4.2. Effect of lexical cue

The fact that listeners' ratings of the Martian utterances did not show a random distribution implies that it is possible to interpret paralinguistic meaning only from prosody without any lexical cues. However, complicated interactions between the STIMULUS TYPE and the TONE TYPE can be seen in Figures 5 and 6, particularly for the Korean utterances, although details are not discussed in the present paper. For example, the Korean sentence whose basic meaning entails annoyance in the declarative, [hatsima] ('don't do it'), received the lowest valence score, and showed less variation in score compared to other types of stimulus. Therefore, the same prosody would signal different paralinguistic meaning depending on the lexical content of a given utterance. A strict and straightforward mapping between semantics and prosody must not be assumed.

4.3. Duration and pitch

For both emotional dimensions, pitch range expansion tended to be associated with intensification, i.e. a greater degree of positivity and excitement. Lengthening was associated with low valence (negativity) but high arousal (excitement).

4.4. Boundary tones

The results generally support the observations that, other things being equal, the complex boundary tones are likely to signal negative emotions and a high degree of arousal, such as irritation or annoyance. However, the effects of lexical cue and

variations in duration and pitch span suggest that it would be counterproductive to assume a straightforward mapping between the type of boundary tone and its basic meaning, whether it is linguistic or paralinguistic. In addition, the pitch movement at the boundary and final lengthening would interact with more global characteristics of speech, such as overall pitch register and variations in speaking rate (e.g. [2]).
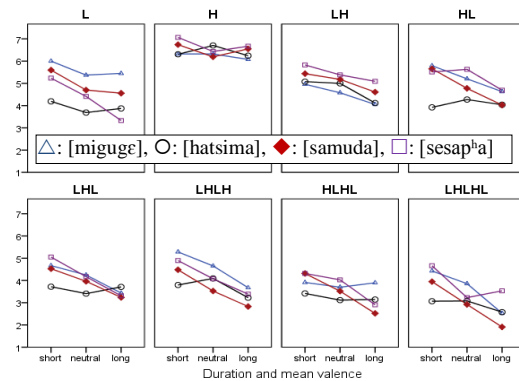
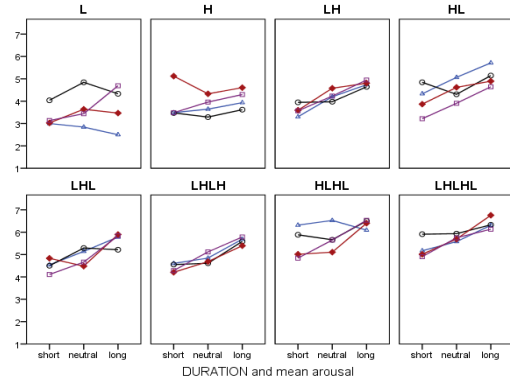Figure 5: *Valence rating for DURATION. Each line represents a different Stimulus Type*

Figure 6: *Arousal rating for DURATION. Each line represents a different Stimulus Type*

5. Conclusions

The emotional interpretation of eight Korean boundary tones did not differ between the two listener groups (native Korean speakers vs. native English speakers learning Korean). The complex boundary tones tended to be perceived as more negative and excited than the simple tones. An increase in pitch span signalled higher positivity and excitement, whilst lengthening signalled higher negativity and excitement. However, the complicated interactions between tone type, utterance-final lengthening, and the lexical cue suggest that the mapping between the type of boundary tone and its paralinguistic meaning is not straightforward.

6. Acknowledgements

7. References

[1] Ladd, D. R. "Intonational Phonology", 2nd ed. Cambridge University Press, Cambridge, UK, 2008.

[2] Pell, M. D., Paulmann, S., Dara, C., Alasseri, A. & Kotz, S. A. "Factors in the recognition of vocally expressed emotions: A comparison of four languages", Journal of Phonetics, 37: 417-435, 2009.

[3] Grabe, E. & Post, B. "Intonational variation in English", Speech Prosody, 343-346, 2002.

[4] Chen, A., Gussenhoven, C. & Rietveld, A. "Language-specificity in perception of paralinguistic intonational meaning", Language and Speech, 47: 311-350, 2004.

[5] Chen, A. "Perception of paralinguistic intonational meaning in a second language", Language Learning, 59: 367-409, 2009.

[6] Jun, S.-A., "K-ToBI (Korean ToBI) labelling conventions (Ver. 3.1)", UCLA Working Papers in Phonetics, 99: 149-173, 2000.

[7] Jeon, H.-S., "Prosodic Phrasing in Seoul Korean: The Role of Pitch and Timing Cues", Unpublished PhD thesis, University of Cambridge, 2011.

[8] Lee, H.-Y., "Kwukeumsenghak" (Korean Phonetics), Thayhaksa, Seoul [in Korean], 1996.

[9] Kim, H. R. S. "A high boundary tone as a resource for a social action: The Korean sentence-ender –ta", Journal of Pragmatics, 42: 3055-3077, 2010.

[10] Park, M.-j. "The Meaning of Korean Prosodic Boundary Tones", Brill, Leiden, the Netherlands, 2012.

[11] Boersma, P. & Weenink, D. "Praat: Doing Phonetics By Computer" [Computer program]. Version 5.3.23, 2012.

[12] Cho, H. S. "Etude des propriétés acoustiques de la structure prosodique du coréen", unpublished PhD thesis, Université Aix-Marseille I. 2009.

[13] Bradley, M. M. & Lang, P. J. "Measuring emotion: the self-assessment manikin and the semantic differential", Journal of Behavior Therapy and Experimental Psychiatry, 25: 49-59, 1994.