DISCRIMINATION OF ENGLISH INTONATION CONTOURS BY NATIVE SPEAKERS AND SECOND LANGUAGE LEARNERS

Hyekyung Hwang, Amy J. Schafer, & Victoria B. Anderson

Department of Linguistics
University of Hawai'i, Hawai'i
{hyekyung; aschafer; vanderso}@hawaii.edu

ABSTRACT

Previous work has shown that advanced Korean learners of English (L2ers) are less effective than native English speakers (L1ers) at using English intermediate phrases (ips) to establish syntactic boundaries [11]. This study investigated whether the effect is due to perceptual differences between L1ers and L2ers, based on the interplay between phonology and perception (e.g., [8, 9]). L1ers and L2ers listened to pairs of phrases in an AX task that crossed boundary strength with intonational contour. Little variation was found between L1ers' and L2ers' discrimination patterns, which correlated highly with each other. Both groups were more sensitive to falling vs. level contour contrasts than rising vs. level contrasts (in the context tested) and were more responsive to contrasts in contour than in boundary strength**.** The results suggest that the L2ers' poor use of ips in comprehension likely rests primarily on difficulty with prosody-syntax mappings.

Keywords: intonation, perception, *Fø* contour, boundary strength

1. INTRODUCTION

Previous studies of cross-linguistic speech perception argue that phonology can shape perception. For example, Hume et al. [9] demonstrated differences in the use of place of articulation cues in stop transitions in native speakers of American English versus Korean, consistent with differences in the languages' phonological contrasts, and Huang [8] found closer perceptual distance for Mandarin native speakers than for English native speakers for two Mandarin tones that participate in a tone sandhi process.

Grabe et al. [6] examined effects of native language experience in the perception of intonation contours in Southern British English intonation phrases. Native speakers of Southern British English, Iberian Spanish, and Mandarin Chinese heard rising vs. falling intonation contours and rated the degree of difference in the *Fø* movements of stimuli pairs. The results consistently showed similar falling/rising differentiation and highly correlated perceptual configurations among the three language groups, which Grabe et al. argued was due to a universal auditory mechanism. However, reliable differences appeared in the perceptual organization of stimuli within each of the rising and falling groups between speakers of different languages. The authors interpreted these findings as showing language-specific perceptual configurations modified from the outputs of a universal auditory mechanism by speakers' native language experience.

The present study addresses the same issue – whether the native phonological system can influence the perception of non-native contrasts in prosodic phrasing – by comparing the perception of similarities and differences among intonational contours between American English native speakers and Korean learners of English. The phonological inventory of prosodic phrasing contours differs between Korean and English. Phonologically, Korean accentual phrases (APs) require final rises (although this varies at the phonetic level and is overridden by intonation phrase boundary tones), while English ips can end in falling, rising, or level contours. Thus, some English ip tonal patterns such as falling contours might be difficult for Korean L2ers to perceive and process.

Hwang and Schafer [11] investigated the role of prosodic phrasing in sentence processing and found that advanced Korean L2ers use falling contour ip boundaries markedly less effectively than English L1ers to recover English syntactic structure but performed similarly with intonation phrase boundaries. This study examined whether Korean L2ers' difficulty is due to perceptual differences. Specifically, we investigated how English L1ers and advanced Korean L2ers discriminate falling vs. level contour contrasts and rising vs. level contrasts in English, as well as more generally whether both groups can reliably detect ip boundaries.

2. METHODS

Forty English L1ers and forty advanced Korean L2ers (who had lived in the US) listened to pairs of phrases in an AX discrimination task that crossed boundary strength (same vs. different size) with contour: rising

(L*L*H-) vs. level (L*L*L- or L*L*) or falling (H*H*L-) vs. level (H*H*H- or H*H*). Each pair was played once with a 200ms inter-stimulus interval between tokens in the pair. Then, participants chose between two options, "same" or "different." Judgment decisions and reaction times were collected.

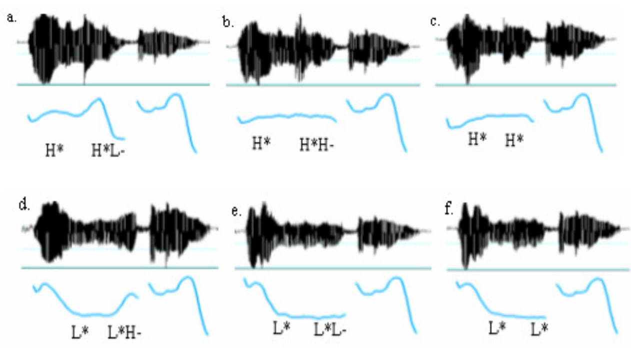
2.1. Materials

The phrase *Meringue and melon from Manila* was produced with six types of prosody (see example (1) and Fig. 1) in four tokens per contour type by a female native speaker of American English trained in phonetics and ToBI. The materials were recorded into 16-bit digital sound files sampled at 22.5 kHz. The segmental information of the entire phrase was kept constant across all items. The phrase *from Manila* was always realized with H*L-L% tones. The first NP *Meringue* ended with a word-level boundary. The end of the second NP *melon* was marked by an ip boundary with a phrase tone, either H- or L-, plus phrase final lengthening, or by a word-level boundary. The first two NPs *Meringue* and *melon* were associated with the same choice of pitch accent, i.e., both had either H* or L*.

(1) Meringue and melon from Manila

a.	H*	H*L-	H*L-L%
b.	H*	H*H-	H*L-L%
c.	H*	H*	H*L-L%
d.	L*	L*H-	H*L-L%
e.	L*	L*L-	H*L-L%
f.	L*	L*	H*L-L%

Figure 1: Sample waveforms and *Fø* tracks.



To minimize potential differences in the production of the phrase *from Manila*, utterances were spliced at the onset of the preposition *from*. The beginning of each token was spliced with a single token of the PP whose pitch range was most comparable with most of the beginning fragments. Before pairing prosodic contours, 16 English native speakers judged the appropriateness of each phrase's pronunciation. These judgments confirmed that the pronunciations were comparably acceptable across the six prosodic contours.

In addition, analyses of duration and *Fø* for the materials verified that the test phrases employed the intended prosodic contours. First, final-syllable lengthening marked ip boundaries. Paired t-tests on the duration of *melon* indicated that *melon* delimited by ip boundaries was significantly longer than *melon* demarked by word-level boundaries (p's< .01). As expected, paired t-tests on the duration of *Meringue* did not find any significant differences (p's> .159). Multiple *Fø* measurements were compared for *Meringue* and *melon*, including the *Fø* for the onset of *Meringue* and *melon*, the *Fø* maximum at any point in the stressed vowel of the two NPs, the *Fø* minimum for the stressed vowel, and the *Fø* at the end of *melon*. These verified each tonal contrast (see [10] for further details).

The six types of prosodic contours were then combined (see Fig. 2) to create four types of experimental, four types of control, and six types of filler trials, repeated in four blocks. Experimental and control trials paired different contours; filler trials paired identical tokens. Presentation order was randomized.

Each contour type was tested with four different tokens to generalize minor pronunciation variation for the contour type. To keep the experiment a reasonable length, not all combinations of tokens were presented. Instead, each block presented two of the four tokens for each prosodic contour type, with the four tokens of each contour distributed across blocks. Each contour pair was presented in both orders (e.g., L*L*H- vs. L*L*L-, and L*L*L- vs. L*L*H-), but because this was not counter-balanced across tokens, order was not treated as a factor in the design. Altogether, each of the four experimental pairing types was presented a total of eight times to each listener (2 orders x 4 blocks), in four distinct pairings of tokens.

2.2. Predictions

The advanced Korean L2ers were predicted to discriminate a rising vs. level contrast (pairs 1, 3 in Fig. 2) more easily than a falling vs. level one (pairs 2, 4) if the final rise of the Korean AP boundary influences their perception of English phrase tones. Alternatively, if their perception is not influenced by the Korean AP boundary tone, there should not be such an effect of rising vs. falling contours. English native speakers should detect both contrasts well.

In addition, pairs with differences in boundary strength (pairs 3, 4) should be more saliently distinct than pairs without such contrasts (pairs 1, 2). If both

Korean L2ers and English L1ers can perceive the presence vs. absence of ip boundaries, their performance should show a significant effect of boundary strength.

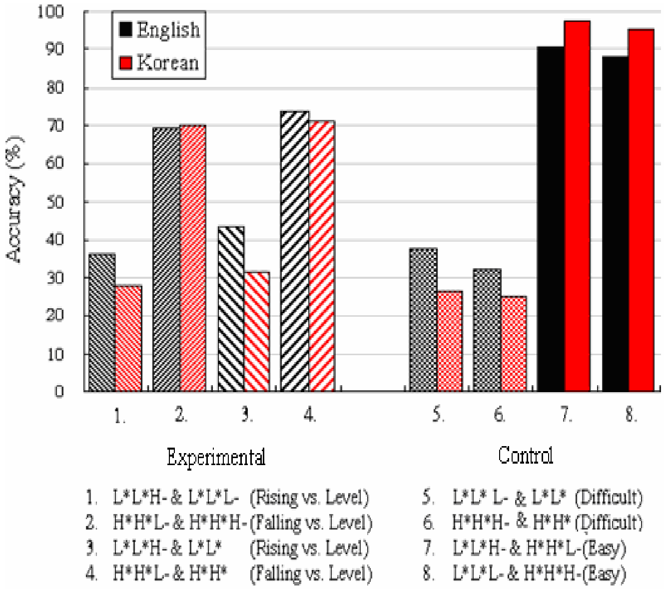
The control pairs [L*L*L- vs. L*L*] and [H*H*H- vs. H*H*] were expected to be hard to discriminate because they are distinguished primarily by subtle durational cues (i.e., phrase final lengthening); these pairs also provided a test of ip detection. The pairs [L*L*L- vs. H*H*H-] and [L*L*H- vs. H*H*L-] were predicted to be easy for L1ers and L2ers because of the large pitch distinction available between the two contours in each pair.

3. RESULTS

The results for the control materials (pairs 5–8) matched the predictions, ensuring that the task was appropriate. A repeated measures ANOVA performed on the accuracy data found a significant main effect of “Difficulty” (F(1, 78)=1063.7, p<.01) but no main effect of “Language group” (F(1, 78)=0.2, p< .634). There was however a significant interaction of “Difficulty” with “Language group” (F(1, 78)=18.4, p< .01). Unpaired t-tests verified that L2ers showed significantly higher accuracy for contour contrasts (“easy” controls) than L1ers (p< .01), but significantly lower accuracy for boundary strength contrasts (“difficult” controls) (p< .011). Fillers were very accurately categorized (93.66% correct).

Overall, participants had difficulty discriminating experimental pairs and showed an apparent bias for “same” responses; in particular, rising vs. level contour contrasts showed below-chance-level performance (39.85% correctly categorized as “different” for English L1ers and 29.69% for Korean L2ers). Results of a 2 (contour contrast) x 2 (boundary strength contrast) x 2 (language group) repeated measures ANOVA for judgment accuracy on the four experimental pairs showed main effects of both “contour contrast” (F(1, 78)=253.5, p< .01) and “boundary strength contrast” (F(1, 78)=8.0, p< .01); performance was better with falling-level pairs and for pairs with contrasting boundary strengths. The interaction of “boundary strength contrast” with “language group” (F(1, 78)=1.3, p< .251) was not significant, but the interaction of “contour contrast” with “language group” was, (F(1, 78)=4.2, p< .05). As shown in Fig. 2, the contour effect was larger for the L2ers than the L1ers. Contra predictions, the L2ers performed similarly to the L1ers with falling-level discriminations, but less well with rising-level ones. Finally, no significant main effect of “language group” was found (F(1, 78)=2.1, p< .154).

Figure 2: Percentage of correct judgments for each type of experimental and control pair in English L1ers and Korean L2ers.



Paired t-tests within language groups revealed that both language groups were significantly more accurate in discriminating the falling-level contrast than the rising-level contrast (all p’s< .01). The presence/absence of ip boundaries was somewhat more salient for English L1ers than Korean L2ers. The effect of an ip was marginal for English L1ers (p< .071 for rising-level contours with versus without an ip and p< .096 for falling-level contours) while they were non-significant for Korean L2ers (p< .266 and p< .562 respectively).

In sum, there was little variation in the discrimination of English prosodic contours between English L1ers and Korean L2ers. Although there was some indication of better performance on boundary strength contrasts by L1ers and pitch accent contrasts (easy controls) by L2ers, L2ers' discrimination accuracy correlated highly with L1ers' (r= .966 based on 16 experimental and 16 control pairs). That is, 93.4 percent of the variance in Korean L2ers’ performance can be accounted for by the English L1ers’ discrimination patterns, suggesting systematic correspondence for most tokens in the two language groups’ perception. The accuracy results were corroborated by the judgment times. E.g., the Korean L2ers’ judgment times for experimental and control pairs were highly correlated with those from English L1ers (p< .01, r= .823).

4. DISCUSSION

The present study provides evidence that the perception of prosody is largely consistent between native speakers and advanced second language learners. For both groups the falling-level contrast was more salient than the rising-level contrast and contour contrasts were more salient than boundary strength contrasts. One possible explanation for the first effect is the frequency of tonal contrasts. The H*H*(L-) patterns are more frequent in English than the L*L*(H-) ones (e.g., [5]). Since the Korean L2ers were advanced learners and immersed in English-speaking countries, they may have become sensitive to these relative frequencies, and thus both groups may have performed better with the more frequent forms. A second possibility is the salience of the tonal contrasts. The size of the tonal rise in the L*L*H- tunes (mean = 63.54Hz) was smaller than the size of the tonal drop in H*H*L- tunes (mean = 120.04Hz) (see also [2, 12]). Further, with a following H* on *Manila*, H*H*L- was more distinct from H*H*H- than L*L*H- was from L*L*L- (see Fig. 1). Future work with a following low tone (i.e., an L* on *Manila*) would help to separate these accounts.

Regarding boundary strength, duration contrasts did not contribute to discrimination as effectively as contour contrasts, even for English L1ers. Wightman et al. [13], Chavarría et al. [4], and Gussenhoven and Rietveld [7] studied acoustic correlates of prosodic phrase boundaries, and found a high correlation between perception of the prosodic phrase boundary level and phrase final syllable lengthening in English. There has been no previous work, to our knowledge, that has contrasted the perception of ip boundaries in falling, rising and level end contours. However, Beach [1] showed a trading relations effect on the identification of English prosodic boundaries. The current results are consistent with her findings: the durational effect of the ip boundary was highly comparable across the stimuli, and so discrimination was better in pairs with greater differences in pitch movement in the boundary region. Because the boundary strength contrasts in the materials were subtle and only occurred in 29% of the trials, it is also possible that listeners' attention was drawn more to tonal contrasts than durational contrasts.

The high correlation between the two language groups in this study is consistent with a universal perception system. However, unlike [6], advanced L2ers participated in this study. Therefore, the current results might also be due to re-shaping of perceptual strategies during learning, or transfer of knowledge from L1 to L2. Further research with participants at different proficiencies is needed to determine the source of the similar perceptual patterns found here.

Finally, the study suggests that the perception of prosody is different from the use of prosody for higher level linguistic analyses. That is, the poor use reported by [11] of ip boundaries (in H*L- contours) for L2ers' syntactic parsing was not caused by markedly different perceptual abilities. We conclude that the effect was likely due to higher level processes, such as categorizing a phonetic pattern as an ip or forming prosody-syntax mappings. This follows from learning models in which learning begins with perceptual processes and progresses to higher-level abstract categories, e.g., [3].

5. REFERENCES

[1] Beach, C.M. 1991. The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *J. of Memory and Lg.* 30, 644-663.

[2] Beckman, M.E., Ayers, G. 1997. *Guidelines for ToBI labeling*. Ms. Columbus, OH: Ohio State University.

[3] Carroll, S.E. 2002. Induction in a modular learner. *Second Language Research 18,* 224-249.

[4] Chavarría, S., Yoon, T-J., Cole. J., Hasegawa-Johnson, M. 2004. Acoustic Differentiation of ip and IP Boundary Level: Comparison of L- and L-L% in the Switchboard corpus. *Proc. 2nd Speech Prosody* Nara, 333-336.

[5] Dainora, A. 2001. *An empirically based probabilistic model of intonation in American English*. Ph.D. dissertation, U of Chicago.

[6] Grabe, E., Rosner, B.S., Garcia-Albea, J.E., Zhou, X. 2003. Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech* 46, 375-401.

[7] Gussenhoven, C., Rietveld, A.C. 1992. Intonation contours, prosodic structure and preboundary lengthening. *J. of Phonetics* 20, 283-303.

[8] Huang, T. 2001. The interplay of perception and phonology in Tone 3 Sandhi in Chinese Putonghua. *OSU Working Papers in Linguistics* 55, 23-42.

[9] Hume, E., Johnson, K., Seo, M., Tserdanelis, G., Winters, S. 1999. A cross linguistic study of stop place perception. *Proc. 14th ICPhS* San Francisco, 2069-2072.

[10] Hwang, H. 2007. *Prosodic phrasing in sentence comprehension: Evidence from native English speakers and native Korean-speaking second language learners of English*. Ph.D. dissertation, U of Hawai'i.

[11] Hwang, H., Schafer, A.J. 2006. Native speakers and second language learners use fine-grained prosodic distinctions in parsing early closure sentences. Poster presentation at the 19th Annual CUNY Conference of Human Sentence Processing, NY, NY.

[12] Pierrehumbert, J. 1980. *The Phonetics and Phonology of English Intonation*. Ph.D. dissertation, MIT.

[13] Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., Price, P.J. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *JASA* 91, 1707-1717.