Perception of lexical pitch accent and rising shape in South Kyungsang Korean

Hyunjung Joo[1] & Mariapaola D'Imperio[2]

[1]Rutgers University, [2]Aix-Marseille University & Laboratoire Parole et Langage, CNRS
hyunjung.joo@rutgers.edu, mariapaola.dimperio@univ-amu.fr

ABSTRACT

This study investigates how South Kyungsang Korean (SKK) listeners perceive lexical pitch accents depending on three factors: f_0 rise shape, f_0 peak alignment, and segmental duration. Specifically, a two-alternative forced choice task was conducted to see how they categorize different f_0 contours as either H or LH for phrase-medial monosyllabic words. Results showed that f_0 rise shape and segmental duration influenced SKK listeners' categorization, while no effect of peak alignment was observed. Interestingly, they responded to more convex shapes as LH, while more concave shapes, as H. Moreover, shorter duration induced an H categorization, while longer duration was related to LH. Results suggest that SKK listeners use both f_0 shape and segmental duration as important cues for tonal contrast, though f_0 shape shows stronger categorical effect than duration. Thus, f_0 shape information is important to determine phonological representation of pitch accents, as opposed to strict tonal alignment in Autosegmental-Metrical theory.

Keywords: perception, lexical pitch accent, f_0 rise shape, South Kyungsang Korean

1. INTRODUCTION

A great attention has been paid on how to define phonological units of continuous f_0 contour in Intonational Phonology. Broadly, two types of approach have been proposed. The first approach is a configurational approach, where the accentual f_0 contour is considered as a whole, such as rises and falls [5, 11, 22]. That is, f_0 shape in rises and falls would directly reflect intonational categories. The other approach is the Autosegmental Metrical (AM) theory, focusing on local f_0 turning points (targets), such as High and Low [3, 14, 17]. In this approach, a contour is defined with sparse f_0 turning points, which are connected to each other by linear interpolation. However, AM turning points (targets) might not be necessarily enough to characterize an f_0 contour, since they cannot capture f_0 shape differences by limiting to target interpolation [9, 15, 20].

Previous studies showed that several factors such as either f_0 peak alignment or f_0 rise (and fall) shape may play a crucial role in perceiving sentential meaning of f_0 contour in some post-lexical pitch accent languages. For example, in earlier studies, American English listeners have been shown to categorically perceive L*+H (uncertainty) vs. L+H* (assertion) pitch accents depending on the f_0 peak timing [18]. Specifically, they identified earlier peaks as L+H* and later peaks as L*+H. Recently, the Tonal Center of Gravity (TCoG) approach has been proposed to account for both f_0 peak alignment and f_0 contour shape [1, 2]. TCoG calculates the weighted location of the area under the curve of f_0 rising (or falling), being a perceptually important reference.

In fact, recent studies [1, 2] showed that f_0 interpolation shape may influence American English listeners' categorization of L+H* vs. L*+H. That is, listeners responded to more scooped rise shapes as L*+H, while more domed shapes were categorized as L+H*. Moreover, a similar effect was obtained by modifying fall shape within a rise-fall [2]. In similar vein, in French, the rise shape of the LH* accent, being either more concave or more convex, influence the perceptual shift of modality from continuation to question [10]. Specifically, more concave shapes were identified as question, while more convex shapes as continuation.

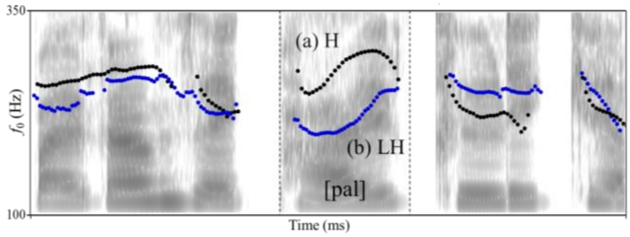

Figure 1: f_0 contour of (a) 'foot' (/pal/ with H tone) and (b) 'shade' (/pal/ with LH tone), embedded in the phrase-medial position.

Unlike pitch accent languages, South Kyungsang Korean (SKK) is a lexical pitch accent language, where f_0 contour is used to distinguish homophone pairs [6, 8, 13, 16]. For monosyllabic words, there are two types of tonal contrast: H and LH[1]. For example, /pal/ with an H tone means 'foot', while /pal/ with a

[1] Several studies used two tonal types, H and LH (or Rising), for monosyllabic words in SKK [6, 8, 13, 16]. Specifically, H tone was described as having a quite higher initial f_0, followed by a low f_0, with a shorter duration. LH tone was described as having a slightly lower initial f_0 and a later f_0 peak alignment (or a low rising contour), with a longer duration.

LH tone means ‘shade.’ As shown in Fig.1, we can notice that the overall rise shape for H and LH appears to be quite different, and in addition the f_0 peak for H appears to be aligned slightly earlier than the LH peak.

First, note that f_0 height appears to be much higher for H than for LH.[2] Crucially, for the H tone, f_0 rises fast with an increasing slope, showing a concave shape. On the other hand, though both tones are L at the beginning, for the LH tone f_0 curves first downward and then rises with a shallow slope, leading towards a convex rising shape. This may suggest that f_0 rise shape can be an important cue for SKK listeners to differentiate two homophonous lexical items contrasting in lexical pitch accent.

A previous study investigated the effect of factors such as f_0 peak alignment and segmental duration, using only one lexical item in an isolated context [7]. One of the findings was that earlier peaks and shorter duration were identified as H, while later peaks and longer duration were identified as LH. Here, we hypothesize that f_0 rise shape might also affect the H vs. LH contrast in SKK. Therefore, the present study aims to examine how SKK listeners perceive their lexical pitch accent, H vs. LH, depending on three factors: f_0 peak alignment, segmental duration, and f_0 rise shape (concave vs. convex).

2. METHODS

2.1. Stimuli

Target words consisted of three monosyllabic homophone pairs (*gan* [kan] (‘taste’ and ‘liver’), *bam* [pam] (‘night’ and ‘chestnut’) *bal* [pal] (‘foot’ and ‘shade’)), contrasting by lexical pitch accent (H vs. LH) as shown in Table 1. Test words were embedded in a carrier sentence, *yobeon-eneun mauseulo kkamansaeg* ___ *keullighaela* [jobʌnænɯn mausɯɾo k'amansæk ___ kɯlɾikhæɾa] ‘Click black ___ with the mouse this time’, in an IP-medial position. Target words were always preceded by the word *kkamansaeg* [k'amansæk] (‘black’) to derive the preceding context as HL. 15 repetitions of these sentences were recorded by a female SKK speaker in a sound-proof booth in Seoul, South Korea, using a SHURE KSM44A microphone. In order to clearly induce the tonal contrast, she was asked to assign a contrastive focus[3] on the test words with a normal speech rate. Prosodic rendition was checked by a trained ToBI transcriber.

Word	Tone type	Meaning
/kan/	H	taste
	LH	liver
/pam/	H	night
	LH	chestnut
/pal/	H	foot
	LH	shade

Table 1: Three monosyllabic homophone pairs.

2.1.1. *Stimulus resynthesis*

Tonal alignment and shape, as well as the duration of the target word were manipulated in Praat [4]. First, the manipulation of f_0 rise shape is illustrated in Fig. 2. As for the target word, the f_0 rise started from the vowel onset (190 Hz), and it ended at the offset of the sonorant coda (290 Hz). In order to create convex and concave f_0 rise shapes, the f_0 value at the midpoint of the rime was increased by 10 Hz steps from the lowest to the highest f_0 level, resulting in 11 steps. The f_0 value of the preceding HL tone ended at the same level (190 Hz) as the vowel onset of the target word. Segmental duration was ambiguous, with mean duration of all H and LH words (290 ms for /kan/, 252 ms for /pal/, and 280 ms for /pam/).

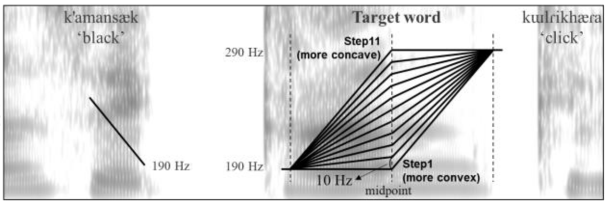

Figure 2: Manipulation of f_0 rise shape.

Second, peak alignment manipulation is illustrated in Fig. 3. Peak alignment was varied from 10% to 100% of the rime, resulting in 10 steps. Each alignment step was increased with 10% of the rime duration (25 ms for /kan/, 23 ms for /pal/, 24 ms for /pam/). The f_0 value of the preceding HL tone ended with the same f_0 value (215 Hz) as at the target word vowel onset. Segmental duration was ambiguous, with the mean duration of all H and LH words (290 ms for /kan/, 252 ms for /pal/, and 280 ms for /pam/), as well as f_0 values for both the L and the H targets (L = 215 Hz; H = 265 Hz).

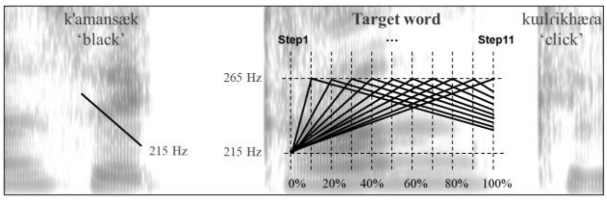

Figure 3: Manipulation of peak alignment.

[2] We haven't factored in f_0 height difference for the categorization of H vs. LH, which will be investigated in a follow-up study. In Fig.1, mean f_0 minimum value was 239 Hz for H and 189 Hz for LH; mean f_0 maximum value (i.e., f_0 peak) was 290 Hz for H and 234 Hz for LH.

[3] In South Kyungsang Korean, only one pitch accent survives in an accentual phrase (AP), while the other pitch accents in the AP become deaccented [13]. Hence, we assigned (a contrastive) focus onto the target word in the AP-medial position to see the lexical pitch accent clearly.

Last, as for the adjustment of segmental duration (Fig. 4), mean duration from the onset of stop release to the offset of sonorant coda was increased by 10 ms, resulting in 11 steps of a continuum. f_0 values of the target word were set to be ambiguous, being the average of all H and LH words (L = 215 Hz; H = 265 Hz). As in the alignment manipulation, the preceding HL tone ended with the same f_0 value (215 Hz) at the target word vowel onset.

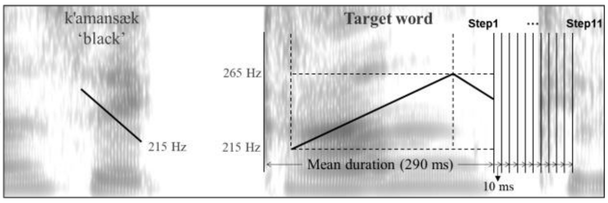

Figure 4: Manipulation of duration.

2.2. Participants

Twenty-five native listeners of South Kyungsang Korean (15 females, 10 males) participated in this study. They were aged from 18 to 27 (mean = 21.6), born and raised in the southern part of the Kyungsang region, mainly Pusan and neighboring cities in South Korea. None of them had hearing problems and they were compensated for their participation.

2.3. Procedure

Participants saw visual stimuli for each target word through a laptop and heard sound stimuli using a headphone (Sony MDR-7509) in a sound-proof booth in Seoul and Ulsan, South Korea. A two-alternative forced choice experiment was carried out using Psychopy [19]. Before the experimental session, the participants had a short training session to become familiarized with the experimental settings. In each trial, they were asked to look at the visual stimuli on the screen upon hearing the sound stimuli and responded by pressing the button of the visual target that matched the target word. Visual stimuli were presented with two different images for a homophone pair (e.g., for the word /pam/, 'night' on the left and 'chestnut' on the right) on the screen. A total of 384 stimuli (3 factors (11 f_0 rise shapes + 10 f_0 peak alignments + 11 durations) x 3 items (/kan/ for 'taste' and 'liver', /pam/ for 'night' and 'chestnut', /pal/ for 'foot' and 'shade') x 4 blocks) were presented. The four blocks were presented in a randomized order, with an interval between the blocks. Among the 25 subjects, the result of one participant was discarded due to overall wrong responses.

2.4. Measurement and statistical analysis

In order to see how SKK listeners perceive f_0 contours depending on rise shape, peak alignment, and segmental duration, response percentage for each step in the continuums were measured. Then, a series of mixed-effects logistic regression models were carried out by using glm package in R [21]. To examine the effects of the three factors on tone identification, H vs. LH response (H coded as 1, LH coded as 0) was used as a dependent variable, while continuum steps were used as a fixed factor. All fixed factors were included as continuous variables. Random intercepts and slopes were set for Subject and Item and *p*-values less than .05 were considered to be significant.

3. RESULTS

3.1. Rise shape

Results showed to what extent SKK listeners categorically perceived H vs. LH depending on f_0 rise shape, either concave or convex, in the f_0 continuum, as shown in Fig. 5. H response percentage for f_0 rise shape showed a curvature pattern similar to a sigmoid function curve. That is, the H response percentage gradually increased from Step 1 to Step 11, which found to be significant in the binomial regression analysis (b = 0.47, z = 7.93, p <.001). The parameter estimates of the model indicated that one step increase in the continuum for Rise shape yielded a change in the log odds of selecting H tone by 0.47, with a positive difference of 19% in the probability of selecting H tone. The category boundary (50% crossover point) was at step 7.03, which was calculated from the simple glmer function (intercept: -2.788; coefficient of Step: 0.397), as shown in the vertical line of Fig. 5. Note that below Step 7 listeners responded as LH (with the rise being more convex), while above Step 7 they responded as H (with the rise being more concave). At the two extremes of the continuum, listeners responded to Step 1 mainly as an LH tone (H response percentage: 0.12), whereas they responded to Step 11 mainly as an H tone (H response percentage: 0.8).

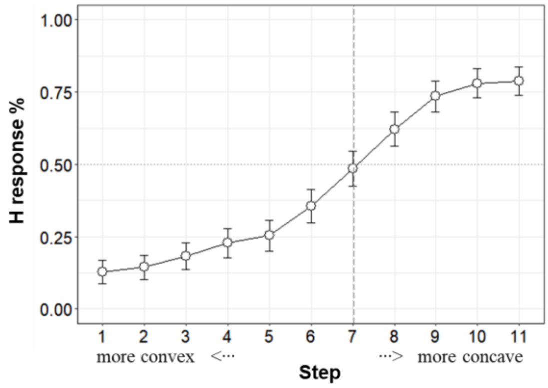

Figure 5: H response percentage for f_0 rise shape.

3.2. Peak alignment

Results showed that SKK listeners did not differentiate H vs. LH depending on f_0 peak alignment, with either an earlier or a later peak as shown in Fig. 6. Even though the H response percentage for f_0 peak alignment appeared to decrease from earlier to later peaks (b = -0.03, z = -7.4, p <.001), H response percentage for all steps were above chance. That is, regardless of the timing of f_0 peak, they perceived all steps in the continuum mainly as H, though more H responses were obtained with earlier peaks.

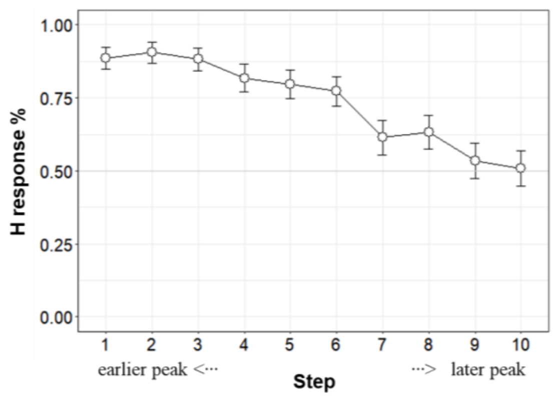

Figure 6: H response percentage for f_0 peak alignment.

3.3. Segmental duration

Results show to what extent SKK listeners categorize H vs. LH depending on segmental duration, either shorter or longer, as shown in Fig. 7. Note that the H response percentage for segmental duration gradually decreased from Step 1 to Step 11 in the continuum, showing a statistical significance of the Step factor ($b = -0.02$, $z = -6.81$, $p < .001$). The parameter estimates of the model indicated that one step increase in the continuum for Duration yielded a change in the log odds of selecting H tone by -0.02, with a negative difference of 5% in the probability of selecting H tone. Specifically, for the two extremes of the continuum, listeners responded to Step 1 (shorter duration) mainly as H tone (H response %: 0.73), whereas they responded to Step 11 (longer duration) mainly as LH tone (H response %: 0.29). The category boundary was at step 4.54, according to the glmer function (intercept: 0.77; coefficient of Step: -0.017), as shown by the vertical line in Fig.7.

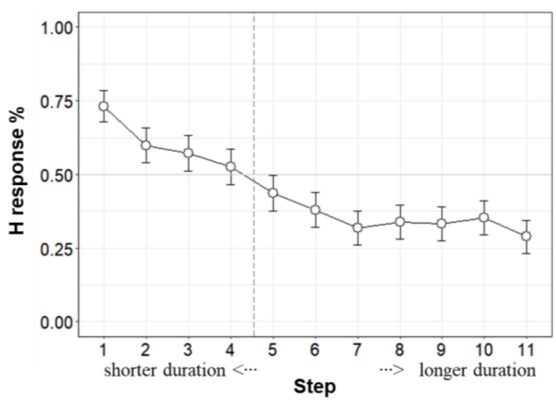

Figure 7: H response percentage for segmental duration.

4. DISCUSSION

This study showed how SKK listeners perceive H vs. LH lexical pitch accents mainly depending on two factors: f_0 rise shape and segmental duration. Crucially, H response percentage showed an effect of both f_0 rise shape and segmental duration. As for the f_0 shape, the H response percentage showed a near-sigmoid curve pattern, with a category boundary between H and LH at Step 7. Note that SKK listeners responded to more convex shape as LH tone while more concave shape as H tone, in an ambiguous duration and f_0 height target. This is consistent with some pitch accent languages in the literature, where tune perception differed depending on the shape of the nuclear pitch accent [1, 2, 10]. This is also in line with the TCoG hypothesis, claiming that the rise (and fall) shape is perceptually crucial to determine pitch accent category in American English [1, 2].

As for the segmental duration, the H response percentage gradually decreased with durational increase, with a category boundary at step 4.54. Specifically, SKK listeners responded to shorter duration as H, while longer duration as LH. This is consistent with previous finding that longer duration increases the identification of an LH tone in SKK [7]. This further confirms the fact that the LH contour tone requires longer duration to be better identified [23]. However, there was no effect of f_0 peak alignment, so that SKK listeners did not differentiate H vs. LH depending on the f_0 peak timing. This is contrary to the previous finding that later peaks were identified as LH tone in SKK, even when the stimuli were presented in isolation [7].

Taken together, both f_0 rise shape and segmental duration influenced tonal perception in SKK. In fact, the overall distribution of the H response was larger for f_0 rise shape (0.12-0.8) than for segmental duration (0.29-0.73), showing a stronger categorical effect. However, further studies are needed to see how these factors interact with each other and which cue is more robust for lexical pitch accent perception in SKK.

Hence, results showed that SKK listeners are mainly sensitive to rise contour shape, in line with what proposed in the configurational approach and TCoG [1, 2, 5, 11, 22] for lexical pitch accents. This also shows that fine-phonetic detail about f_0 shape may be stored in the phonological representation of the pitch accents [1, 2, 12]. However, the absence of a significant effect of f_0 peak alignment supports that it is not sufficient to characterize tunes as a sequence of tone targets. Hence, the target-and-interpolation within AM theory seems rather problematic given the impact of f_0 shape between tone targets [3, 14, 17].

5. CONCLUSION

Our study showed that both f_0 rise shape and segmental duration play a crucial role for SKK listeners to differentiate homophone pairs carrying contrastive lexical pitch accents, i.e. H vs. LH. Specifically, more convex shaped rises as well as longer target words led to a higher LH perception, while more concave shaped and shorter items were mainly identified as H. On the other hand, tonal alignment did not affect lexical accent identification. A potential impact of additional factors such as target f_0 height and its interactions with rise shape and segmental duration need to be examined in further studies.

6. REFERENCES

[1] Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*, 3(2), 337-383.

[2] Barnes, J., Brugos, A., Veilleux, N., & Shattuck-Hufnagel, S. 2021. On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity. *Journal of Phonetics*, 85, 101020.

[3] Beckman, M. E., & Pierrehumbert, J. B. 1986. Intonational structure in Japanese and English. *Phonology*, 3, 255-309.

[4] Boersma, P., & Weenink, D. 2009. Praat: doing phonetics by computer (Version 5.1. 05) [Computer program]. Retrieved May 1, 2009.

[5] Bolinger, D. L. 1951. Intonation: levels versus configurations. *Word*, 7(3), 199-210.

[6] Chang, S. E. 2007. *The phonetics and phonology of South Kyungsang Korean tones.* Unpublished doctoral dissertation, University of Texas, Austin.

[7] Chang, S. E. 2013. Effects of fundamental frequency and duration variation on the perception of South Kyungsang Korean tones. *Language and speech*, 56(2), 211-228.

[8] Do, Y. A., & Kenstowicz, M. 2010. A note on phonological phrasing in South Kyungsang. *Seoul International Conference on Linguistics (SICOL) 2010, June, 23-25, 2010, Korea University, Seoul.*

[9] D'Imperio, M. 2000. *The role of perception in defining tonal targets and their alignment.* Unpublished doctoral dissertation thesis, Ohio State University.

[10] Dorokhova, L., & D'Imperio, M. 2019. Rise dynamics determines tune perception in French: The case of questions and continuations. *In International Congres of Phonetic Science ICPhS2019*.

[11] Hirst, D., & Di Cristo, A. 1998. *A survey of intonation systems*. Intonation systems: A survey of twenty languages, 144.

[12] Kimball, A. E., & Cole, J. 2016. Pitch contour shape matters in memory. In *Proceedings of the International Conference on Speech Prosody*, 8, 1171-75.

[13] Kim, J., & Jun, S. 2009. Prosodic Structure and Focus Prosody of South Kyungsang Korean. *Language Research*, 45.1, 43-66.

[14] Ladd, D. R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[15] Ladd, D. R., & Schepman, A. 2003. "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of phonetics*, 31(1), 81-112.

[16] Lee, H., & Zhang, J. 2014. The nominal pitch accent system of South Kyungsang Korean. *Journal of East Asian Linguistics*, 23(1), 71-111.

[17] Pierrehumbert, J. B. 1980. *The phonology and phonetics of English intonation.* Doctoral dissertation, Massachusetts Institute of Technology.

[18] Pierrehumbert, J. B., & Steele, S. 1987. How many rise-fall-rise contours. *In Proceedings of the XIth International Congress of Phonetic Sciences*.

[19] Peirce J. W. 2009. Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2(10), 1-8.

[20] Pitrelli, J. F., Beckman, M. E., & Hirschberg, J. 1994. Evaluation of prosodic transcription labeling reliability in the tobi framework. *In Icslp*, 1, 123-126.

[21] R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[22] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech communication*, 46(3-4), 220-251.

[23] Zhang, Jie. 2002. *The effects of duration and sonority on contour tone distribution: A typological survey and formal analysis.* New York: Routledge.