The Second International Conference on Tone and Intonation
18-20 November 2023, Singapore, Singapore

No effects of f0 manipulation and phrase position in Korean word recognition

Constantijn Kaland[1], *Matthew Gordon*[2], *Jiyoung Jang*[3], *Argyro Katsika*[2]

[1]Institute of Linguistics, University of Cologne, Germany
[2]Department of Linguistics, University of California, Santa Barbara, United States
[3]Hanyang Institute for Phonetics and Cognitive Sciences of Language, Seoul, Korea

ckaland, mgordon, jiyoungljang, argyro
[@uni-koeln.de | @ucsb.edu | @hanyang.ac.kr]

Abstract

Work on pitch accent languages has shown that f0 shape facilitates word recognition. That is, listeners are faster recognizing words which are accented than words which are unaccented. Less is known about the role of f0 in languages that do not makes use of pitch accents. In those languages, it could be that f0 still contributes to word recognition to some extent or that word recognition is aided by boundary tones, rather than pitch accents. Korean is a language without pitch accents, making use of boundary tones only (i.e., edge language). The current study extends earlier work by replicating its experimental paradigm, i.e. it tests the effects of f0 shape and phrase position in a word recognition experiment. The results confirm the status of Korean as an edge language and are interpreted in a prosodic typological context.

Index Terms: f0, word recognition, phrase position, Korean, prosodic typology.

1. Introduction

Typologically, languages have been categorised for their prosodic features at the word- and phrase-level, including the rhythmical characteristics that originate from both levels [4]. The most important distinction concerns the one between different types of prosodic prominence: either head, the edge, or head and edge. In short, linguistic constituents are marked prosodically either at their head, at the edge(s) or at both locations. A second prosodic typological distinction is made at the word level, i.e. whether a language has stress, tone, both or none. The idea is that word prosodic marking only concerns heads, and that consequently no edge language has word prosody. See [5] for a typological account of the relationship between word- and phrase-level prosody. In typological studies, Korean is therefore analysed as an edge language, similar to West Greenlandic [6].

Korean is traditionally analysed within the Autosegmental Metrical framework (AM; e.g., [1]) as a language in which prosodic phenomena operate at the level of the Accentual Phrase (AP, e.g. [2]). Although the length of an AP may vary, generally it is reported as being the size of a word (minimally) or a small phrase (e.g., [3]). The edges of Korean APs are marked tonally by a low-high accent (LH) AP-initially and -finally [4]. As for the word level, Korean is analysed as having no word stress or any specific prosodic marking at the lexical level (e.g. [2]). Thus, Korean is further analysed as having no pitch accents, in line with the common assumption in AM-approaches that pitch accents at the phrase level align with syllables that are stressed at the word level.

There is limited work on the role of prosody to word recognition for languages with only phrase-level prosody. This issue is not trivial, in particular because speakers can make changes in prosodic phrasing in order to draw focus on specific words [4]. It is furthermore known from perception studies on head languages that listeners recognize accented words faster than unaccented words [8]. This effect was furthermore found to depend on phrase position, with accents in final phrase position facilitating word recognition to a larger extent than in earlier positions [9]. These results match with those from language acquisition studies showing that in child directed speech, important words are often post-positioned and produced with wider f0 excursions [10]. It is thus not a priori clear to what extent word recognition is affected by f0 shape and phrase position in edge languages, such as Korean.

Previous perception studies on Korean have mainly focused on word *segmentation*, i.e. prosodic cues with which listeners can break down the speech stream into words ([11], [12]). One study used acoustically manipulated stimuli presented as trisyllabic sequences as words in an artificial language [11]. The manipulated cues concerned duration (final lengthening), amplitude (initial strengthening), initial f0 (high) and final f0 (high), with no manipulation of any syllable as a baseline. Listeners' task was to detect whether a sequence was part of an artificial language that they familiarized with prior to the experiment. Duration, amplitude and final f0 were shown to facilitate listeners' segmentation performance. It was argued that despite Korean lacking word prosody, APs often coincide with words and provide therefore helpful prosodic cues to their segmentation.

The role of f0 in Korean word segmentation was also investigated in a word spotting task using manipulated f0 movements in the AP (LH, LL, HH, HL) and before the left edge of the AP (pre-boundary L or H) [12]. Results showed that listeners could segment most accurately with a pre-boundary H, and an initial L tone in the AP (LH or LL). The tonal combination *across* the AP boundary thus matters. A subsequent segmentation experiment tested the role of pre-boundary final lengthening, i.e. before the start of the AP, in addition to the tonal configurations just mentioned. The results showed that final lengthening helped listeners only when the pre-boundary tone is L, which occurs infrequently in Korean. The latter effect thus showed that f0 and duration do not have a cumulative effect. Only when f0 does not provide clear cues listeners do use duration.

The current study on Korean extends previous work on American English and Papuan Malay [13] that the extent to which Papuan Malay has pitch accents, like American English [1]. To this end, a word recognition task was carried out in which listeners responded as quickly as possible which out of two written words on a screen they heard in the stimulus. Stimuli were spoken carrier phrases containing a target word with original or manipulated (flat) f0, and in phrase-medial or phrase-final position. Results showed that in both languages listeners

10.21437/TAI.2023-14

were faster recognizing the target word when it had original f0 (as opposed to a flat f0) and when it occurred in phrase-final position (as opposed to phrase-medial position). Given the similarity between the languages, it seemed that Papuan Malay listeners benefited from accentuation, just like the American English listeners. Despite these similarities, the prosody of both languages is likely to be different, both with respect to the type of word stress (e.g., [14] and the use of phrase accents [15]. These results suggest that Papuan Malay might be a head or head/edge language, whereas American English has been traditionally analysed as a head language. It is therefore important to extend this work by investigating an edge language such as Korean.

It thus remains to be seen whether Korean listeners indeed benefit from f0 shape and phrase position in word recognition. Given the lack of pitch accents and word prosody in Korean we hypothesize that word recognition for Korean listeners would not be facilitated by f0 to the extent that it is for American English and Papuan Malay listeners [13]. There could, however be an effect of phrase position in that phrase-final words are recognized faster, either as a generic recency effect [16] or because it coincides with the right edge of the AP, facilitating segmentation [11]. To investigate these questions, we replicate the task used in [13], of which the methodological details are outlined in the next section.

2. Methodology

A reaction time (RT) experiment was designed in which listeners task was to identify a target word from an auditory stimulus. Two experimental variables were crossed: target words occurred either with their original f0 contour or with a flat (manipulated one, and they either occurred in phrase-medial position or in phrase-final position (2 x 2). Unless otherwise stated, all methodological descriptions match the ones in [13].

2.1. Participants

26 native speakers of Korean (variety) participated in the experiment: 12 F, 14 M, M age: 31.4, age range: 27-39. None of them had hearing problems.

2.2. Stimuli

Stimuli consisted of spoken carrier phrases with the target word embedded medially or finally (see 1). Note that we henceforth refer to phrase-medial and phrase-final, where 'phrase' refers to the respective positions in the carrier phrase, not in the phrase (i.e. IP or AP) in the phonological sense. The carrier phrases were produced by a female native speaker of Korean (Seoul variety) and were chosen such that they matched the phrase constructions used in [13]. All further processing of the recordings was done using Praat [17]. The intensity of the recorded carrier phrases was scaled to minimize acoustic differences between the stimuli that were not part of the experiment. Target words were all disyllabic and had a [CV.CV] syllable structure. For each target word a distractor was chosen. Both target and distractor were presented on screen and participants had to indicate which of them they heard in the stimulus. Distractors had the same initial syllable as the target. In this way, the uniqueness point (UP) at which participants could identify the target occurred in the middle of the word, warranting attention to the relevant cue (f0) in the target's acoustic realisation. The recorded carrier phrases were annotated and segmented for the target word and its two syllables to obtain the UP timestamps for each stimulus. There were 20 medial targets (med) and 20 final targets (fin).

(1) Carrier phrases with the target in phrase-medial (MED) and in phrase-final (FIN) position.

MED	네가	말한	그 단어	나비는,	나는	모르겠어.
	ni-ga	mal.han	kɯ.tan.ʌ	[T]-neun,	na-nɯn	mo.lɯ.ke.sʌ
	you-NOM	mention	the.word	[T]-TOP,	I-TOP	don't know

'the word [T] you mentioned, I don't know'

FIN	나는	모르겠어,	네가	말한	그 단어	나비.
	na-nɯn	mo.lɯ.ke.sʌ,	ni-ga	mal.han	kɯ.tan.ʌ	[T]
	I-TOP	don't know,	you-NOM	mention	the word	[T]

'I don't know the word [T] (you mentioned).'

2.3. F0 manipulation

F0 manipulation was carried out using Time-Domain Pitch-Synchronous Overlap-and-Add resynthesis (TD-PSOLA; [18]). The manipulation concerned flattening of the f0 on the target word such that it followed the natural declination of the carrier phrase (see Figure 1 for examples in all conditions). This was done by stylization of the contour with a resolution of two semitones. All pitch points in the interval of the target word were removed. In most cases, this procedure flattened any f0 movement on the target sufficiently. However, the f0 level at the end of the word was sometimes higher than at the start and sometimes much lower at the end. To obtain a naturally declining f0 over the target word, additional pitch points were removed until reaching the first point that was lower than the last original point (at the start of the medial targets) or by adding a phrase-final point (final targets). Note that the original contour also underwent TD-PSOLA resynthesis (without shape manipulation), to balance out potential effects of the resynthesis on the audio quality. After f0 manipulation, the total number of stimuli was 80 (original: 20 MED, 20 FIN; manipulated: 20 MED, 20 FIN).

2.4. Procedure

The experiment was designed in PsyToolkit ([19], [20]), which provides an online environment to design and run experiments via a web-browser on a PC. The software generates html pages to instruct the participants, present the stimuli and collect the data. Participants first received instructions about the task. Participants completed a practice round consisting of five stimuli to familiarize themselves with the task. At the end of the practice round participants were asked whether they felt they needed to practice more or whether they were ready to start the actual task. When more practice was needed, participants were presented additional stimuli. In the actual task, the target and distractor words were written on the screen, either at the left or right side (randomly assigned). Participants' task was to indicate as fast as possible which of the two words on the screen they heard in the phrase. They could indicate the word by pressing '1' (left word) or '0' (right word) on their keyboard (see Figure 2). The key presses were saved and participants needed to press the space bar to continue to the next stimulus. This was done as self-paced RT experiments have been shown to lead to lower rates of missed responses and to improve participants'

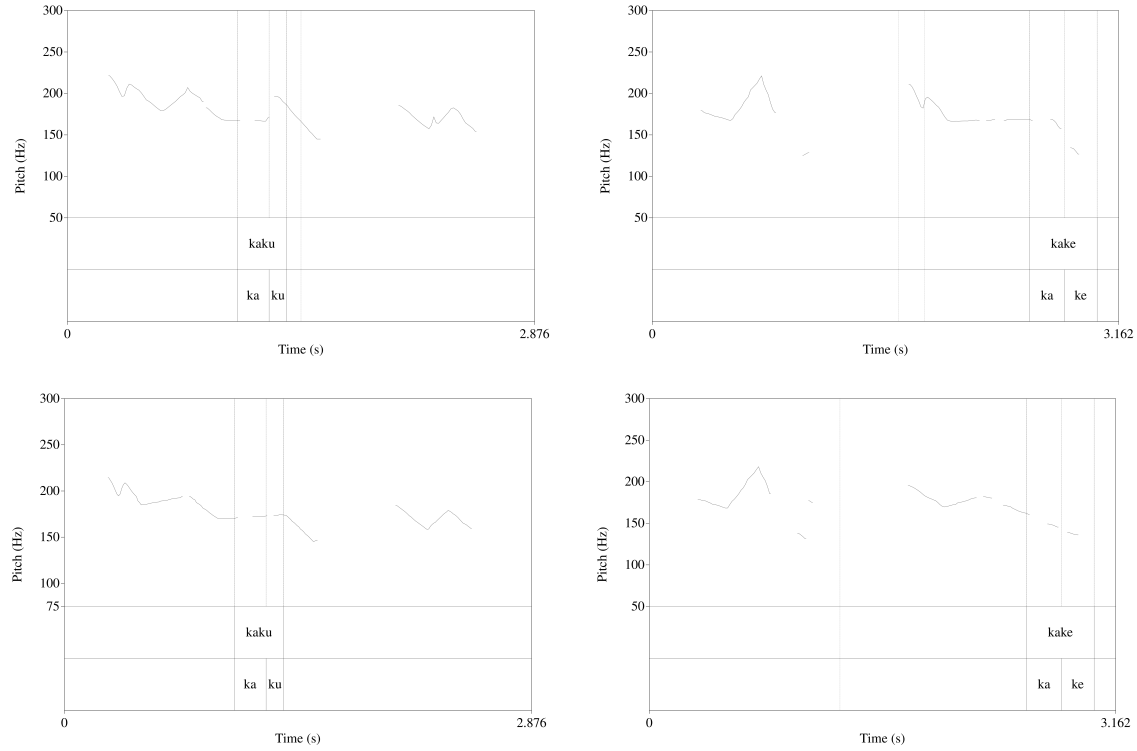

Figure 1: *Examples of target words with original (top) and manipulated (bottom) f0 contour in medial (left) and final (right) phrase position. Tiers show segmentation on the word (top) and syllable (bottom) level.*

compliance [21]. Responses were saved as either 'correct' or 'incorrect' including the RT as measured from the UP.

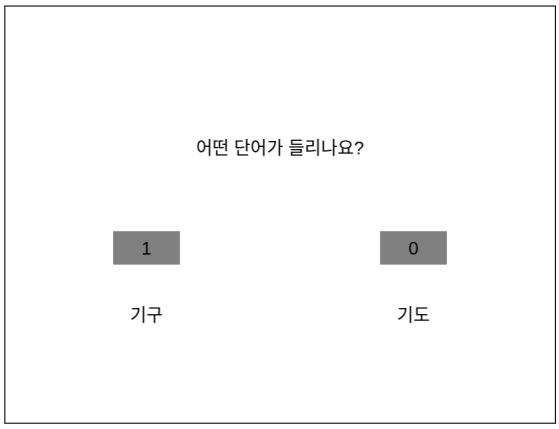

Figure 2: *Reaction times measured from target word onset (top) and offset (bottom) in phrase-medial and phrase-final position when the cue syllable was the first (grey) or second (white).*

2.5. Statistical analysis

RTs longer than 2 seconds (N = 37) and RTs shorter than 200 ms (N = 8) were discarded, as they were considered unreliable for further analysis. The RTs were log-transformed to obtain a normal distribution of the residuals. Thereafter, outliers were removed from the data, following a procedure outlined in [22]. This involved removal of data points with absolute standardized residuals exceeding two standard deviations. The number of segments before the UP was calculated for each stimulus in order to take this into account as a factor potentially affecting the RTs. In addition, the trial number was taken into account, referring to the position in the experiment at which the stimulus was presented. The latter was done to take potential fatigue effects on the RTs into account.

Linear mixed modelling was performed in R ([23], [24]) using the `lmerTest` package [25] on the log-RTs as measured from the UP, with the interaction of f0 shape (original, manipulated) and phrase position (medial, final), and with number of segments before UP (two, three), trial number (range 1-80) as fixed factors. Participant and item were included as random intercepts.

3. Results

Table 1: *Mean reaction times (RT) in milliseconds and as logarithm in each experimental condition. Standard deviations between brackets.*

Position	f0 shape	RT (ms)	log(RT)
med	original	627.68 (174.75)	6.40 (0.28)
	manipulated	629.38 (183.85)	6.40 (0.29)
fin	original	638.45 (195.01)	6.41 (0.30)
	manipulated	627.66 (179.89)	6.40 (0.29)

The mean RTs show little to no difference between the experimental conditions (Table 1). Thus, Korean listeners took ap-

proximately 630 ms from the UP to recognize the target word, regardless of its f0 shape and regardless of its phrase position. The results of the LMM (Table 2) indicate no significant effect of any of these factors. In addition, there was no effect of the number of segments before the UP, nor of the presentation order if the stimuli.

Table 2: *Results of the LMM on the log-RTs.*

Factor	b	SE	df	t	p
(Intercept)	6.38	0.07	59.07	94.30	< 0.001
f0 shape	-0.00	0.01	1869.92	-0.19	n.s.
phr.position	0.01	0.04	39.73	0.33	n.s.
segm.b.UP	0.01	0.02	36.27	0.40	n.s.
stim.no.	0.00	0.00	36.28	0.49	n.s.
f0:pos	-0.01	0.02	1853.70	-0.63	n.s.

4. Discussion

This study has found no effects on word recognition latencies in Korean, either by f0 shape or by phrase position. The results seem to indicate that the two factors are not taken into account by listeners in online word recognition. This contrast with listeners of American English and Papuan Malay [13], who where both shown to be faster for words with their original (unmanipulated) f0 contour and when these occurred in phrase-final position. As for f0, the lack of effect on Korean listeners is not surprising. Korean lacks word prosody [4], so listeners apparently do not need prosodic cues to recognize them. The effects on word segmentation found for Korean were in fact originating from a prosodic unit larger than the word (the AP). The results are thus likely to be compatible with work showing that APs aligning with word boundaries may facilitate word segmentation and could speed up recognition [11].

It should also be noted that the degree of f0 manipulation in Korean was less than in previous languages [13], as there were smaller, more subtle f0 movements. For example, phrase-final LL (Figure 1) plausibly did not change sufficiently to be interpreted as a different tone sequence. That is, flattening the f0 still allowed for an LL perception and interpretation. Phrase-medially, the LH tone-sequence was indeed eliminated by the f0 flattening. However, the particle 는 (-neun), which was added to avoid the target word occurring directly before the phrase-break (at the comma), was not manipulated for f0. Given that this particle attaches to the target word, unlike 'there' in American English and 'itu' in Papuan Malay. That is, in the latter two language those are separate words (cf. [13]). The original L on the Korean particle could therefore have facilitated word recognition to some extent.

As for the lack of effect of phrase position, the results are somewhat counterintuitive. The literature has shown that phrase-final positions are privileged in speech processing, as listeners benefit from final lengthening, recency, and a pause after the phrase-final word (e.g. [26]; [27]). Although the phrase-final effect has been partially attributed to predictability, note that that in the current study and in [13] listeners could not make predictions about upcoming words, as the carrier phrase did not give any information about them. Given the effects of phrase position in American English and Papuan Malay in an identical task, predictability is therefore excluded as a possible explanation for the lack of this effect in Korean. It should be noted that final-lengthening is not consistently used by Korean speakers and listeners (see discussion in [27]). Taken together, it cannot be fully excluded that Korean participants in the current study benefited from the phrase-medial particle '-neun' and did not benefit as much from final lengthening phrase-finally. These two factors could have contributed in such a way that no effect of phrase position was found.

From a typological perspective, the f0 is consistent with the analysis of Korean as an edge language. In retrospect, this also indicates that Papuan Malay is more similar to American English than to Korean in terms of word recognition. This would imply a head/edge or head language analysis of Papuan Malay. The results of the current study also indicate that word recognition is just one way of investigating prosodic typological differences. To solidify typological accounts, more crosslinguistic research is needed using identical experimental paradigms.

5. Acknowledgements

Research for this paper was funded by the German Research Foundation (DFG) – Project-ID 281511265 (SFB-1252 "Prominence in language").

6. References

[1] J. Pierrehumbert, "The phonology and phonetics of English intonation," Thesis, Massachusetts Institute of Technology, 1980, accepted: 2009-01-23T14:36:47Z.

[2] S.-A. Jun, "The phonetics and phonology of korean prosody: Intonational phonology and prosodic structure," PhD thesis, The Ohio State University, Columbus, USA, 1996.

[3] S.-A. Jun and J. Fletcher, "Methodology of studying intonation: from data collection to data analysis," in *Prosodic Typology II*, S.-A. Jun, Ed. Oxford University Press, Jan. 2014, pp. 493–519.

[4] S.-A. Jun, "Prosodic typology: by prominence type, word prosody, and macro-rhythm *," in *Prosodic Typology II*, S.-A. Jun, Ed. Oxford University Press, Jan. 2014, pp. 520–539.

[5] M. K. Gordon, "Disentangling stress and pitch-accent: a typology of prominence at different prosodic levels," in *Word Stress*, H. van der Hulst, Ed. Cambridge, UK: Cambridge University Press, 2014, pp. 83–118.

[6] A. Arnhold, "Prosodic structure and focus realization in West Greenlandic," in *Prosodic Typology II*, S.-A. Jun, Ed. Oxford: Oxford University Press, Jan. 2014, pp. 216–251.

[7] Y. Igarashi, "Typology of intonational phrasing in Japanese dialects," in *Prosodic Typology II*, 1st ed., S.-A. Jun, Ed. Oxford University PressOxford, Jan. 2014, pp. 464–492.

[8] A. Cutler and D. J. Foss, "On the Role of Sentence Stress in Sentence Processing," *Language and Speech*, vol. 20, no. 1, pp. 1–10, Jan. 1977.

[9] G. Mehta and A. Cutler, "Detection of Target Phonemes in Spontaneous and Read Speech," *Language and Speech*, vol. 31, no. 2, pp. 135–156, Apr. 1988.

[10] A. Fernald and C. Mazzie, "Prosody and focus in speech to infants and adults." *Developmental Psychology*, vol. 27, no. 2, pp. 209–221, 1991.

[11] S. Kim, "The role of prosodic cues in word segmentation of Korean," in *Interspeech 2004*. ISCA, Oct. 2004, pp. 3005–3008.

[12] S. Kim and T. Cho, "The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean," *The Journal of the Acoustical Society of America*, vol. 125, no. 5, p. 3373, 2009.

[13] C. Kaland and M. K. Gordon, "The role of f0 shape and phrasal position in Papuan Malay and American English word identification," *Phonetica*, vol. 79, no. 3, pp. 219–245, Jun. 2022.

[14] C. Kaland, "The perception of word stress cues in Papuan Malay: a typological perspective and experimental investigation." *Laboratory Phonology*, vol. 12, no. 1, Oct. 2021.

[15] C. Kaland and S. Baumann, "Demarcating and highlighting in Papuan Malay phrase prosody," *The Journal of the Acoustical Society of America*, vol. 147, no. 4, pp. 2974–2988, Apr. 2020.

[16] J. L. Shields, A. McHugh, and J. G. Martin, "Reaction time to phoneme targets as a function of rhythmic cues in continuous speech." *Journal of Experimental Psychology*, vol. 102, no. 2, pp. 250–255, 1974.

[17] P. Boersma and D. Weenink, "Praat: doing Phonetics by Computer," 2022.

[18] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, vol. 9, no. 5, pp. 453–467, Dec. 1990.

[19] G. Stoet, "PsyToolkit: A software package for programming psychological experiments using Linux," *Behavior Research Methods*, vol. 42, no. 4, pp. 1096–1104, Nov. 2010.

[20] ——, "PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments," *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, Jan. 2017.

[21] H. Krinzinger, J. W. Koten, J. Hennemann, A. Schueppen, K. Sahr, D. Arndt, K. Konrad, and K. Willmes, "Sensitivity, Reproducibility, and Reliability of Self-Paced Versus Fixed Stimulus Presentation in an fMRI Study on Exact, Non-Symbolic Arithmetic in Typically Developing Children Aged Between 6 and 12 Years," *Developmental Neuropsychology*, vol. 36, no. 6, pp. 721–740, Aug. 2011.

[22] H. R. Baayen and P. Milin, "Analyzing reaction times," *International Journal of Psychological Research*, vol. 3, no. 2, pp. 12–28, Dec. 2010.

[23] R Core Team, "R: the R project for statistical computing," 2022.

[24] R Studio Team, "RStudio: Integrated Development for R," 2022.

[25] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *Journal of Statistical Software*, vol. 82, no. 13, 2017.

[26] D. J. Foss, "Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times," *Journal of Verbal Learning and Verbal Behavior*, vol. 8, no. 4, pp. 457–462, Aug. 1969.

[27] H.-S. Jeon and A. Arvaniti, "Effects of rhythm and phrase-final lengthening on word-spotting in Korean," *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4251–4263, Jun. 2017.