Compression and truncation: The case of Seoul Korean accentual phrase [*]

Hyesun Cho
(Dankook University)

Edward Flemming
(Massachusetts Institute of Technology)

Cho, Hyesun and Edward Flemming. 2015. Compression and truncation: The case of Seoul Korean accentual phrase. *Studies in Phonetics, Phonology and Morphology* 21.2. 359-382. Compression and truncation are two general strategies adopted by languages in realizing tonal melodies in the face of time pressure. Compression involves adjustment of the phonetic details of tone realization to fit a tone sequence in the time available, while truncation involves deletion of a phonological tone, reducing a melody to fit the segmental material with which it must be associated. Seoul Korean appears to show truncation of tones in response to time pressure resulting from fast speech rate: the Accentual Phrase (AP) is canonically marked by a rise-fall-rise melody (LHLH), but at fast speech rates, APs can be realized with a rising F0 contour. Categorical tone deletion conditioned by speech rate would be theoretically significant because it would imply that a phonological operation can be conditioned by utterance-specific phonetic detail. However, analysis of the realization of the AP melody across a range of speech rates provides evidence that the apparent deletion of tones is actually the end result of compressing the final HLH sequence into such a short interval that the low tone is completely undershot and the two high tones are effectively merged. This analysis illustrates the general point that it is not possible to determine what phonological representation gave rise to a particular phonetic form without explicit analysis of the process of phonetic implementation. **(Dankook University and Massachusetts Institute of Technology)**

Keywords: compression, truncation, Seoul Korean, weighted constraints

1. Introduction

It is a basic property of intonational tunes that they can be associated with a wide variety of segmental material. For example, the English rise-fall-rise tune, L+H* L-H%, can be associated with phrases that range in length from multiple syllables down to one, as in (1a) vs. (1b) (Ladd 2008: 181). Associating a melody like this, which consists of multiple tones, to a single syllable is liable to subject it to time pressure. That is, the melody has to be realized in a short duration, and this can be problematic because there are limits to how fast pitch movements can be realized (e.g. Xu and Sun 2002).

(1) a.   L*+H L   H%   b. L*+H L H%

A driving instructor?   Sue?

It has been observed that languages adopt a variety of strategies for dealing with time pressure (e.g. Silverman and Pierrehumbert 1990, Caspers and van Heuven 1993, Prieto 2005, among others). A basic distinction has been drawn between compression, where the complete tune is squeezed into the time available, and truncation, in which one or more tones are deleted (Grice 1995, Ladd 2008: 181ff., Rathcke 2013). The realization of the English L+H* L-H% tune on a monosyllable is an example of compression: the full rise-fall-rise F0 trajectory is realized by virtue of eliminating the L-plateau, lengthening the syllable, and perhaps increasing the rate of the F0 movements and decreasing their magnitudes. Truncation is exemplified by the Hungarian rise-fall question intonation (L* H L%). When this tune is realized on one syllable, the final low boundary tone, L%, is deleted, resulting in a rising contour (Ladd 2008: 182-3). Multiple examples of both responses to time pressure have been reported – see Ladd (2008: 181ff.) for a review.

Truncation due to syllable count is usually analyzed in terms of constraints limiting the number of tones that can be associated with a syllable. In Hungarian, no more than two tones can be associated with a single syllable, so the tri-tonal question melody cannot be realized faithfully on a monosyllabic phrase. In English, on the other hand, such constraints are low-ranked so there is no limit on the number of tones that can be associated with a single syllable.

The distribution of compression and truncation within and across languages is a matter of theoretical interest. In particular, it would have significant implications for the relations between phonetics and phonology if truncation of intonational tunes can occur in response to time pressure that results from fast speech rate. That is, time pressure arises when an intonational tune has to be realized in a short duration. The duration of segmental material depends on the nature of the segmental content and on speech rate. That is, a monosyllabic phrase is shorter than a polysyllabic phrase, other things being equal, but a monosyllabic phrase produced at a slow speech rate can be longer than a polysyllabic phrase spoken quickly. So time pressure depends on speech rate as well as segmental content. However, standard phonological representations abstract away from speech rate variation, so derivation of segmental durations in a particular utterance is analyzed as a matter of phonetic implementation.

Since truncation involves deletion of a tone from the phonological representation, truncation of an intonational melody as a result of high speech rate would imply that a phonological operation can be conditioned by the phonetic details of segmental duration, and thus that the phonology must

have access to information about phonetic implementation. That is, phonology and phonetic implementation are more tightly integrated than is usually assumed (cf. Pierrehumbert 1980, Flemming 2001). If phonology has no access to phonetic implementation, then we would expect that compression would be the only possible response to time pressure resulting from high speech rate. Compression does not necessarily imply any change to the phonological tone sequence since it involves setting the precise timing of F0 events and the rate and magnitude of F0 movements, all of which can generally be analyzed as aspects of the phonetic implementation of a phonological representation.

However, it is not a simple matter to determine whether truncation can be conditioned by speech rate because it is often not simple to distinguish truncation from compression (Ladd 2008: 183). Truncation is often identified based on impressionistic transcription of utterances with the aid of F0 tracks, but identifying the phonological tone sequence that gave rise to a particular F0 contour is not straightforward because the phonetic implementation of tones is non-trivial and language-specific.

In this paper we exemplify this problem with the intonational tune of the Accentual Phrase (AP) in Seoul Korean. The AP lies above the phonological word in the prosodic hierarchy of Korean, but often consists of a single phonological word (a content word plus any clitics or postpositions) (Jun 2000). The canonical tune associated with an AP is illustrated in Figure 1(a). This tune is analyzed as consisting of the tone sequence LHLH (Jun 2000), and these tones are transparently reflected in the F0 contour: each L corresponds to a local minimum and each H corresponds to a local maximum (the final fall is due to the Low tone which begins the following AP).

Figure 1(b) shows the same AP uttered by the same speaker at a fast speech rate. The tune of the AP might be transcribed as LH since it involves a steady rise to a local maximum, followed by a fall to the Low tone that begins the second phrase. We will argue that this AP actually has an LHLH tune, but as a result of time pressure, the second L tone is not realized by a local minimum in the F0 trajectory. This analysis is viable because compression of a tune does not only involve shifting pitch movements closer together and accelerating the rate of F0 transitions, it can also involve reducing the magnitude of pitch movements, undershooting the targets for tone levels (Prieto 1998, Ladd and Schepman 2003, Ladd 2008: 182)[1]. An LHLH tune can be realized as the F0 contour shown in Figure 1(b) if the final HLH tones are extremely crowded together, and the target for the

---

[1] It is a source of some confusion that the terms 'compression' and 'truncation' have also been used to refer to these two strategies for fitting a tune into a short duration (Grabe et al. 2000: 162). In these terms, compression refers to temporal compression of the F0 contour, with minimal reduction in pitch movements, while truncation refers to undershoot of tone targets with minimal change in the slopes of F0 movements. In practice, languages generally use both strategies to varying extents (e.g. Grabe 1998, Grabe et al. 2000). Here we use compression and truncation in the senses defined by Ladd (2008: 181), as described above.

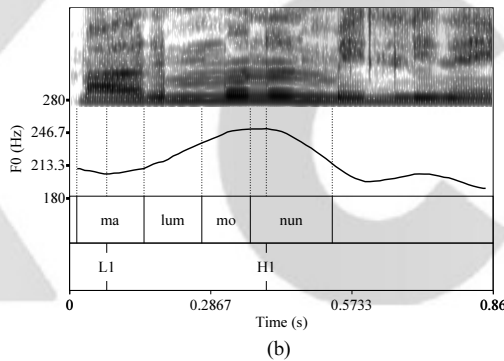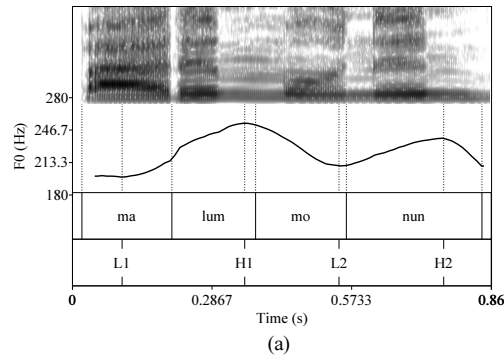second L is undershot so severely that there is no corresponding F0 minimum.

Figure 1. Utterances of an AP by a female speaker (A1) at normal (a) and fast (b) speech rates (*malummonun* [marimmonin] 'A lozenge is').

   Given the extent of possible phonetic variation in the realization of a given tune, impressionistic transcription is insufficient to properly assign a phonological analysis to the tune of a particular utterance. This can only be done in conjunction with an analysis of the mapping from the phonological representation to phonetic realization. We exemplify this approach in a case study of the effects of time pressure on the realization of the Seoul Korean AP melody. This case study illustrates the difficulties of distinguishing truncation from compression, and ways to bring evidence to bear on this question through explicit modeling of the phonetic realization of tones. The results are consistent with the hypothesis that speech rate does not condition deletion of tones in Seoul Korean.

This evidence comes from an experiment eliciting a substantial range of AP durations through variation in segmental content and speech rate. These data allow us to observe and model gradual reduction in the duration of the AP-final HLH sequence as AP duration decreases, with concomitant reduction in F0 lowering for the L tone. The resulting analysis derives the apparent truncation of L as the limit of these compression processes.

A key feature of the analysis is the assumption that compression does not involve uniform shortening or reduction of the F0 contour, rather some properties of the contour are prioritized over others, resulting in greater reduction of certain tones under time pressure. In this case, the duration and magnitude of the initial LH rise is prioritized over realization of the low target for second L tone, with the result that this tone can be completely obscured at fast speech rates. The broader goal of the experiment was to use speech rate variation to probe the nature and relative importance of the various targets that make up the Seoul Korean AP melody (cf. Caspers and van Heuven 1993) – see Cho (2011) for additional results and analysis derived from the same data set.

This paper is organized as follows. Section 2 describes the procedure and results of the experiment on the Seoul Korean Accentual Phrase. Section 3 provides analyses of the results, including a model based on weighted constraints. Section 4 presents general discussion and conclusions.

2. Experiment

2.1 Seoul Korean Accentual Phrase

As mentioned in Section 1, the canonical tune associated with an AP in Seoul Korean is /LHLH/, and we will refer to these tones as L1, H1, L2, H2 respectively (Figure 2). There are two source of variation in this tune, in addition to speech rate. First, if the initial consonant of the AP is tense or aspirated, the initial Low tone is replaced by a High tone, and second, the full tune is only realized if the AP contains four or more syllables. In shorter APs, tones are deleted so there is no more than one tone per syllable. For example, an AP with three syllables typically has an [LH] or [LHH] tonal configuration (Jun 2000, Kim 2013). We will argue below that this pattern is an example of truncation – that is, it involves true tonal deletion as opposed to phonetic compression. However, the experiment reported here includes only four-syllable APs beginning with consonants that are not tense or aspirated.
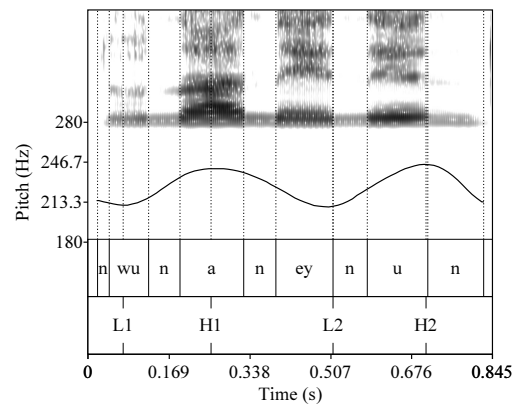
Figure 2. A spectrogram (above) and pitch movement (below) for a female speaker (A1)'s utterance of *nwunaneynun* [nunanenin] 'sister's family is', an Accentual Phrase consisting of four syllables. Normal speech rate. The top text tier shows segmentation, and the bottom text tier shows F0 turning points representing L1, H1, L2, H2 respectively.

2.2 Speech materials and methods

Speech materials consisted of 36 four-syllable APs. Each AP consisted of a three-syllable content word followed by a topic marker nun [nin], e.g. *malineynun* [maɾinenin] 'Mary's family is'. The three syllables in the content words were mostly open syllables (CV) to control for possible effects of syllable structure, but we also included some words that have closed syllables (CVC) in each position, roughly balanced, to give some variation in syllable structure across items. When the final syllable of the content word is a closed syllable, nun becomes an allomorph *un* [in], triggering resyllabification. Therefore to allow closed syllables in this position we used another particle *man* [man] 'only' e.g. *taytongkangman* [tetoŋkaŋman] 'only Taytong River'.

The phrases were followed by a carrier phrase (verb or adjective) with four syllables (See Appendix for the complete list). The consonants in the speech materials were mostly sonorants to facilitate pitch measurement, and a few lenis consonants were included when no real words were available with all sonorants. Tense or aspirated consonants were excluded throughout the speech materials because they are known to affect pitch levels of the following vowels. Vowel height was balanced among high, mid, and low in each syllable, in order to avoid possible effects of intrinsic vowel height (Ohala and Eukel 1987, Whalen and Levitt 1995). In order to prevent speakers falling into artificially rhythmic speech, APs with a different number of syllables (3, 5, 6) were included as filler items (a total of 25 fillers). The order of the sentences was randomized subject to the constraint

that no more than two APs with the same length occurred consecutively. Speakers were first asked to read the sentences naturally, then as fast as possible, and then slowly but naturally. The whole set of sentences was repeated twice at each speech rate.

The subjects were ten native speakers of Seoul Korean, five female (A1-A5) and five male (B1-B5). Eight speakers were in their 20's and 30's (graduate students); one male and one female speaker were in their 40's. None of the speakers reported any speech or hearing problems and they were unaware of the purpose of the experiment. The speech materials were presented on a sheet of paper, and the recording was made in a sound-attenuated recording booth. Speakers used a head-mounted microphone (Shure SM10A) connected to a USBPre sound input device for pre-amplification and digitization. The digitized speech signals were recorded direct to disk using HairerSoft Amadeus II (version 3.8.4) on an iMac, at a sampling rate of 44.1kHz with 16 bits per sample.

The total number of recorded phrases was 2160 (36 targets × 3 speech rates × 2 repetitions × 10 speakers). 11 tokens were discarded because their pitch tracks were not measurable, so a total of 2149 tokens were used for analyses. The recordings were manually labeled for segment boundaries using Praat (Boersma and Weenink 2012). After pitch tracks were smoothed using Praat's smoothing function, pitch targets (L1, H1, L2, H2) were located manually using 'Get minimum pitch' or 'Get maximum pitch' functions in a selected interval in Praat. The L and H tones were located based on F0 minima and maxima. Sometimes there was no F0 minimum corresponding to L2, in which case no measurement was recorded for that tone. After labeling, we collected the timing of each segmental boundary and the timing and scaling of F0 turning points using Praat scripts.

2.3 The results: variant realizations

The instructions to vary speech rate resulted in substantial variation in the duration of APs. In female speakers, APs were on average 1.6 times longer in normal speech than in fast speech, and 1.78 times longer in slow speech than in normal speech. In male speakers, the ratios are 1.6 (normal vs. fast) and 2.28 (slow vs. normal).

The Korean [LHLH] Accentual Phrase shows variant realizations depending on speech rate. Figure 3 shows pitch level values (Hz) for each tone (L1, H1, L2, H2) and the endpoint of the AP (APend). The values are plotted separately for female and male speakers. It can be seen that the AP melody is realized with two clear F0 peaks at normal and slow speech rates (Figure 3 (b) & (c)). Note that the dip (L2) between two H tones is lower in slow speech than in normal speech, for both female and male speakers. On the other hand, there is often no clear dip between H tones in fast speech (Figure 3 (a)), instead a plateau is observed. L2 was hard to measure in such cases. If there was a point that has even a very small pitch decrease between

H1 and H2, that point was taken as L2, but if there was no such point, L2 was left unmeasured.

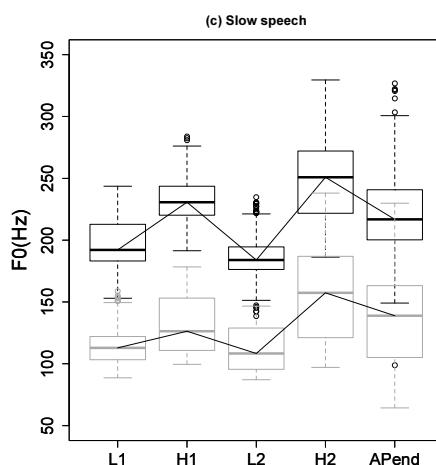
(a) Fast speech
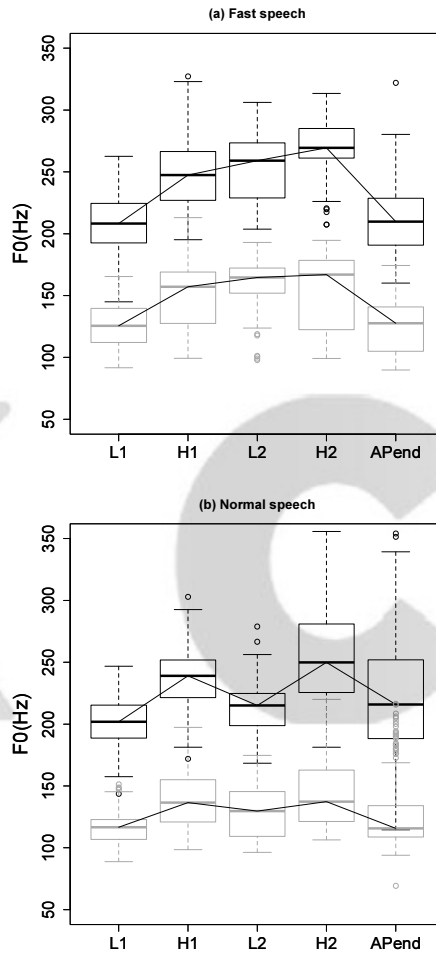
(b) Normal speech

Figure 3. Box plot of the pitch levels (Hz) of each tone (L1, H1, L2, H2) and the endpoint of the AP (APend) for fast (a), normal (b), and slow (c) speech rates. Pitch levels are plotted separately for female (black boxes) and male speakers (grey boxes). The horizontal lines in each box mark the median, upper and lower lines indicate upper and lower quartile, the whiskers outside the box show maximum and minimum, and the dots are outliers.

Figure 4 shows sample pitch tracks of Accentual Phrases produced by a female speaker (A1) (Figure 4(a)) and a male speaker (B1) (Figure 4(b)). It shows the first seven utterances in each speech rate, the first repetition per item. The numbers above each panel are item numbers (as in the Appendix), so each row shows how the same AP is differently realized depending on speech rate.

In the normal speech rate from the female speaker, all four tones are clearly visible. In the examples shown here the two peaks are similar in height. In fast speech, none of the utterances show two peaks, instead a single sharp peak (1), or a rise-fall with a broad peak or a high plateau (5,7) is observed. In slow speech, the most noticeable difference is the higher pitch of H2. This speaker also has a pause between the target AP and carrier phrase (not shown here) in slow speech. These two properties suggest that her target APs are also Intonational Phrases (IPs) in slow speech, with the extra high final peak indicating the presence of an IP-final boundary H%. However, not all speakers show this pattern. Some speakers produced extra high H2 in slow speech without a clear pause following the AP, and others produced a fall in pitch at the end of AP's in all speech rates (like speaker B1, Figure 4(b)). Some within-speaker variation was also observed. In any case, in slow speech all four tones are clearly realized by F0 minima and maxima.
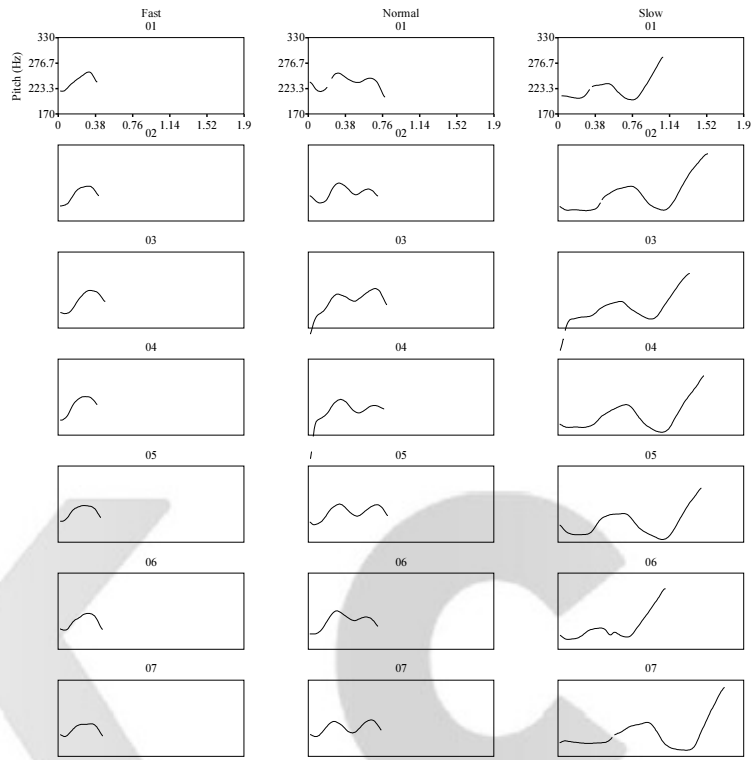
(a) Pitch tracks of speaker A1 (female)
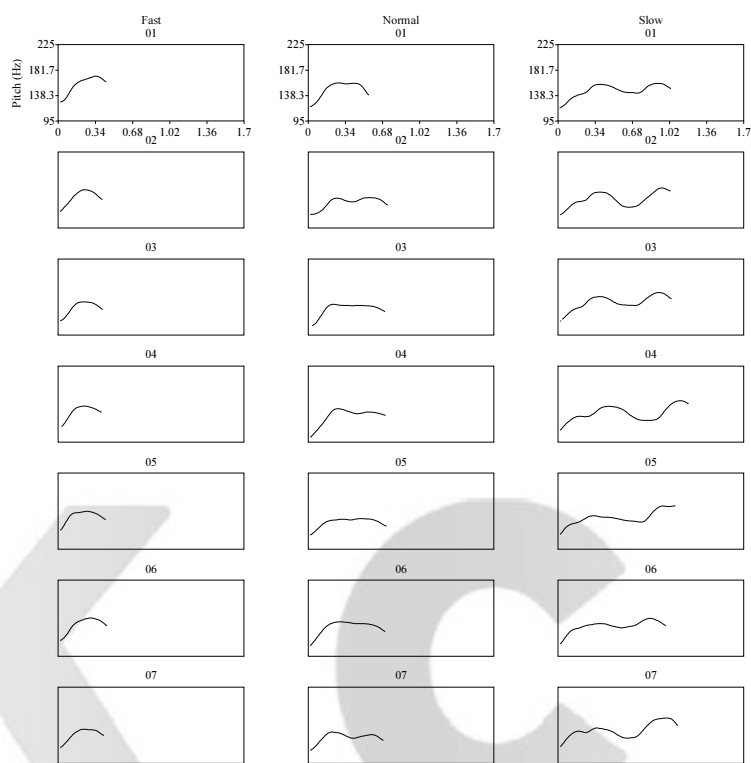
(b) Pitch tracks of speaker B1 (male)

Figure 4. Pitch tracks of the first seven utterances in each speech rate, fast (left column), normal (center column), and slow (right column), produced by a female speaker (A1) (a), and by a male speaker (B1) (b). The beginning and end of a pitch track correspond to the beginning and end of an Accentual Phrase (carrier phrases are not shown). In each panel, the horizontal axis shows time in second, and the vertical axis shows pitch (F0) in Hz. The scales are consistent throughout the panels for each speaker. The number above each panel indicates an item number (see Appendix).

Visual inspection of all the utterances indicates that those AP's which do not end in H% (i.e. those without an extra high H2) can be classified into the four shapes exemplified in Figure 5. Patterns (a) to (c) in Figure 4 differ in their realization of L2 level: in (a) all four tones are fully and clearly realized, in pattern (b) L2 is undershot, so its pitch level is higher than L1, and (c) is realized with a high plateau following the initial rise, with no measurable drop in F0 corresponding to L2.
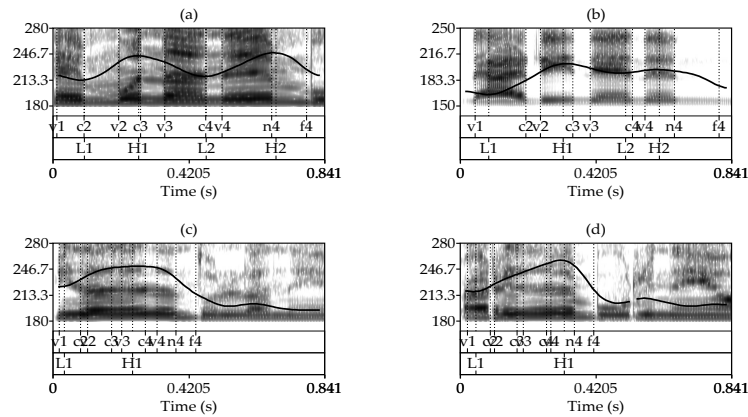
Figure 5. Variants of phonetic realization of four-syllable Accentual Phrase in Korean (reproduced from Cho (2010)). (a) Full scaling (speaker A1, normal rate), (b) undershoot of L2 (speaker A3, normal rate), (c) a rise-fall pattern with a high plateau (speaker A1, fast rate), (d) a rise-fall pattern with a sharp peak (speaker A1, fast rate). The label 'f4' indicates the end of the AP.

These variants are associated with variation in speech rate. In normal and slow speech, all four tones are typically fully realized, as in Figure 5(a) – in particular in slow speech the second L is almost always as low as or even slightly lower than the first L (L1 mean =156Hz, L2 mean=149 Hz, t(674)=13, p<0.0001 (paired t-test)). On the other hand, in normal and fast rates, the L is often undershot as in Figure 5(b). L2 is often much higher than L1, so the L2 is realized merely as a shallow dip between two H peaks (in normal speech, L1 mean=159Hz, L2mean=171 Hz, t(589)=-17, p<0.0001 (paired t-test)). In faster speech, a rise-fall pattern with a high plateau, is often found, as in Figure 5(c), so there is no dip between the two H tones. Finally, only at fast speech rates do we find a pattern such as Figure 5(d), a rise-fall movement with a sharp peak.

Table 1. Number of non-null pitch values depending on speech rate

	Total # of valid tokens	L1	H1	L2	H2
Fast	711	696 97.9%	689 96.9%	68 9.6%	76 10.6%
Normal	720	698 96.9%	709 98.5%	613 85.1%	610 84.7%
Slow	718	683 95%	684 95%	714 99%	718 100%

Table 1 gives preliminary evidence for the relationship between these patterns and speech rate. When there was only one rise-fall pattern, L2 and H2 were not measurable and thus given null values in our data coding. Table 1 shows the number of non-null values for a given tone, depending on speech rate. It can be seen that as speech gets faster, L2 and H2 get harder to locate, i.e. L2 and H2 were observable 9-10% of the time in fast speech, while L1 and H1 were observable close to 100%. In normal speech L2 and H2 were observable in more than 80% of utterances. The null values account for 15%, and arise from plateau patterns, where it is not possible to locate L2 and H2 properly. In slow speech, L2 and H2 were observable almost 100% of the time. In slow speech, on the other hand, non-null values in L1 and H1 are slightly less frequent (95%) than at the other two rates. This comes from the difficulties locating pitch turning points in extremely slow speech. Overall, as speech rate increases L2 and H2 are less likely to correspond to turning points in the F0 contour.

3. Analysis

3.1 Schematic representations and hypotheses

We have now seen in more detail that the second low tone of the AP tune, L2, can apparently be deleted at fast speech rate. However, we have also illustrated that this is only one of several variants of the canonical AP F0 contour where the others involve undershoot of L2 or production of a high plateau with no drop in F0 for L2. All of these patterns can be accounted for in terms of compression processes, as schematized in Figure 6. The full realization of all four tones, at slower speech rates, is illustrated in Figure 6(a). At somewhat faster speech rates, H1 and H2 are realized closer together, as in Figure 6(b), resulting in less time for the fall-rise movement to realize L2, and thus undershoot of this tone target. As speech rate increases further, undershoot of L2 reaches the point where there is no F0 dip between H1 and H2 (Figure 6(c)). At the highest speech rates, H1 and H2 are so close together that they form a single F0 peak (Figure 6(d)). According to this analysis, increased speech rate results in compression of the AP tune only, it does not result in deletion of any of the underlying tones.
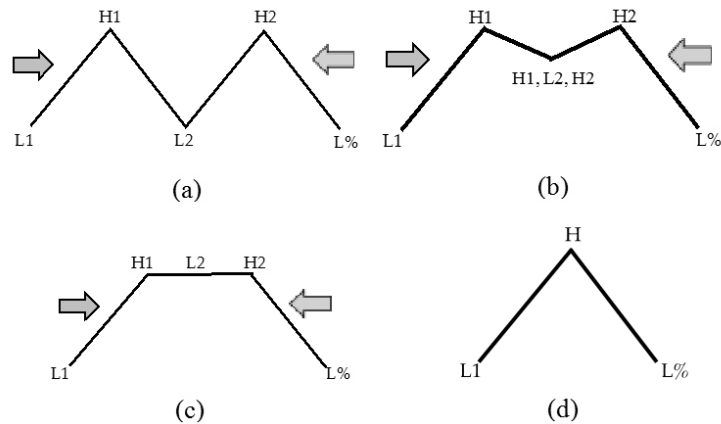
Figure 6. Schematic representations of variant realizations. The arrows at both sides indicate time pressure. Reproduced from Cho (2010).

Our analysis and modeling in the following sections is divided into parts: (i) gradual undershoot of L2 in response to time pressure leads to the formation of a high plateau (changes from (a) to (c) in Figure 6) (Section 3.2) (ii) the high plateau shortens as speech rate increases further (changes from (c) to (d) in Figure 6) (Section 3.3).

3.2 Undershoot of L2

Undershoot of L2 can be understood as avoidance of rapid F0 movements, as illustrated in Figure 7. Figure 7(a) represents an F0 trajectory at a reasonably slow speech rate. In fast speech segmental duration decreases, so if the pitch level of H1 remains relatively stable, reaching the L2 target would result in a very steep fall, as shown by the solid curve in Figure 7(b). Such a rapid movement can be avoided by undershoot of L2 (the dashed line), at the cost of falling short of the target for the level of L2. In other words, there are two conflicting targets involved: the magnitude of the fall (difference between the levels of H1 and L2) and the slope of the fall. It is not possible to satisfy both at the same time under time pressure. If the L2 target is reached, the slope must be steep; if the slope of the fall is kept gentle, then the target level for L2 cannot be reached under time restrictions.
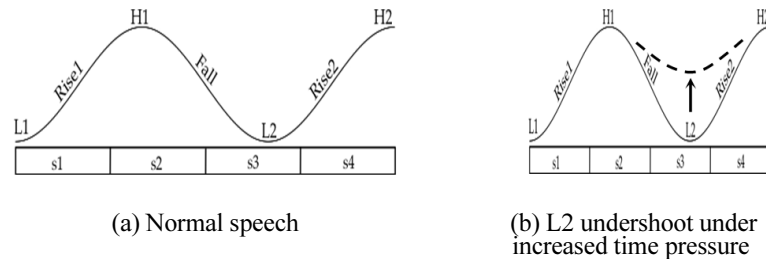
(a) Normal speech	(b) L2 undershoot under increased time pressure

Figure 7. Effects of L2 undershoot (dotted line), if H1 level is maintained.
((a) is reproduced from Cho (2010))

There are various other ways in which a steep fall could be avoided, for example H1 could be aligned earlier in order to leave enough time for the fall, or the level of H1 could be lowered. We have already observed that the level of H1 is not lowered with increasing speech rate – if anything the level of H1 increases in fast speech (Section 2.3, Figure 3, especially in male speakers)). Previous studies have shown that the former strategy is not adopted in Seoul Korean, either. H1 is aligned near the middle of the rime of the second syllable, but at fast speech rates it shifts gradually later relative to this anchor to avoid making the initial rise too short (Cho 2011). On the other hand, L2 is consistently aligned near the middle of the third rime, with very little variation due to speech rate (Cho 2010: 51). Figure 8(a) shows the timing of H1 and L2, together with their segmental anchors ('H1A' the anchor for H1, the middle of the second rime; 'L2A' the anchor for L2, the middle of the third rime). Local speech rate, plotted on the x-axis, is the inverse of the total duration of the second and third syllables, the syllables with which H1 and L2 are associated. As speech rate increases, H1 is delayed relative to its anchor, whereas L2 is not delayed, so the interval between H1 and L2 steadily decreases as speech rate increases. As a result, if the low target for L2 were fully realized, the slope of the fall would be very steep at faster speech rates. Instead L2 is realized higher than its target level. That is, staying close to the target duration for the initial rise, maintaining the alignment target for L2, and avoiding fast pitch movements are prioritized over reaching the F0 target for L2, with the result that this target is undershot under time pressure.
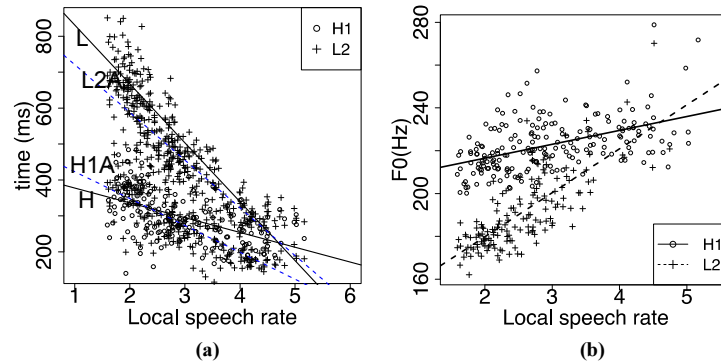
Figure 8. (a) The timing of H1 and L2 for one speaker. The solid lines are regression lines for the timing of H1 and L2 (time is measured from phrase onset). 'H1A' and 'L2A' are the timing of the putative anchor for H1 and the anchor for L2, respectively (dashed lines, H1A=the middle of the second rime, L2A=the middle of the third rime). (b) The pitch level of H1 and L2 for speaker A5 ((b) is reproduced from Cho (2010)).

Figure 8(b) shows the changes in the pitch levels of H1 and L2, depending on speech rate, for one female speaker. L2 is affected more by speech rate than H1 is, conforming to our previous observations. H1 slightly increases in faster speech. The difference between H1 and L2 shows that the magnitude of the fall (H1-L2) decreases as speech rate increases. The level of H1 and L2 is expected to converge at high speech rates, at which point there is no F0 minimum corresponding to L2.

The timing of H1 has been analyzed in terms of a compromise between constraints on the alignment of L1 and H1, and on the duration of the rise between them (Cho 2010). Here we propose to extend this analysis, modeling undershoot of L2 as a function of the duration from H1 to L2 in terms of additional weighted constraints (cf. Flemming 2001, Cho 2010, 2011, Flemming and Cho submitted). As previously mentioned, there are two targets: a target for the magnitude of the fall from H1 to L2 ('M' in Figure 9), and a target for the slope of the transition from H1 to L2 ('S' in Figure 9). 'D' is the duration of the fall from H1 to L2.
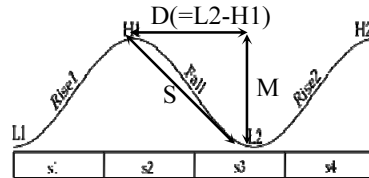
Figure 9. Targets in the falling movement from H1 to L2.
D: duration, S: slope, M: magnitude

These targets are enforced by constraints that penalize deviations from them, as shown in (2). That is, the constraint for magnitude requires the magnitude of the fall, M, to be equal to the target magnitude, T_M and the slope constraint requires the slope, S, of the fall from H1 to L2 to be equal to the target slope, T_S. Deviation from a target incurs a cost equal to the square of the deviation, multiplied by a constraint weight (Flemming 2001). M, S, D are not independent, so one must be calculated from the other two. In (2) M is replaced by S(L-H) (Magnitude=Slope × Duration) for computational simplicity.

(2)   Target   Constraint   Cost of violation
        Magnitude $M=T_M$ $w_M(S(L-H)-T_M)^2$
        Slope $S=T_S$ $w_S(S-T_S)^2$

(M: actual magnitude, S: actual slope, T_M: target magnitude, T_S: target slope, H: timing of H1, L: timing of L2, w_M, w_S: positive weights)

(3) Total cost of violation = $w_M(S(L-H)-T_M)^2 + w_S(S-T_S)^2$

These are conflicting constraints because, given a particular duration D from H1 to L2, achieving a fall with magnitude T_M implies that the slope must be T_M/D, and this will not generally be equal to the slope target T_S. So it is not generally possible to realize both slope and magnitude targets. The actual slope and magnitude are selected so as to minimize the weighted sum of the violations of these constraints (3). The minimum of the cost function in (3) is found by taking the first derivative of the cost function and finding where this derivative is equal to zero. In this way we can derive an expression relating optimal fall magnitude (M) to Slope (S) and fall Duration (D), as in (4). The expression (4) implies that fall magnitude should approach 0 at very short durations and target magnitude (T_M) at very long durations.

(4) $$M = \frac{w_M D^2 T_M + w_S D T_S}{w_M D^2 + w_S}$$

We fitted the expression in (4) to the data as a non-linear mixed effects model using the *nlme* function in the *nlme* package (Pinheiro et al. 2015) in

R (R Core Team 2013). The constraint weight w_M was set to 1 without loss of generality since only the ratios of the constraint weights affect model predictions. The data used were those in which neither H1 nor L2 is null (1338 utterances in total). The model included random effects by speaker for each model parameter, to allow for any variation between subjects in parameter values.
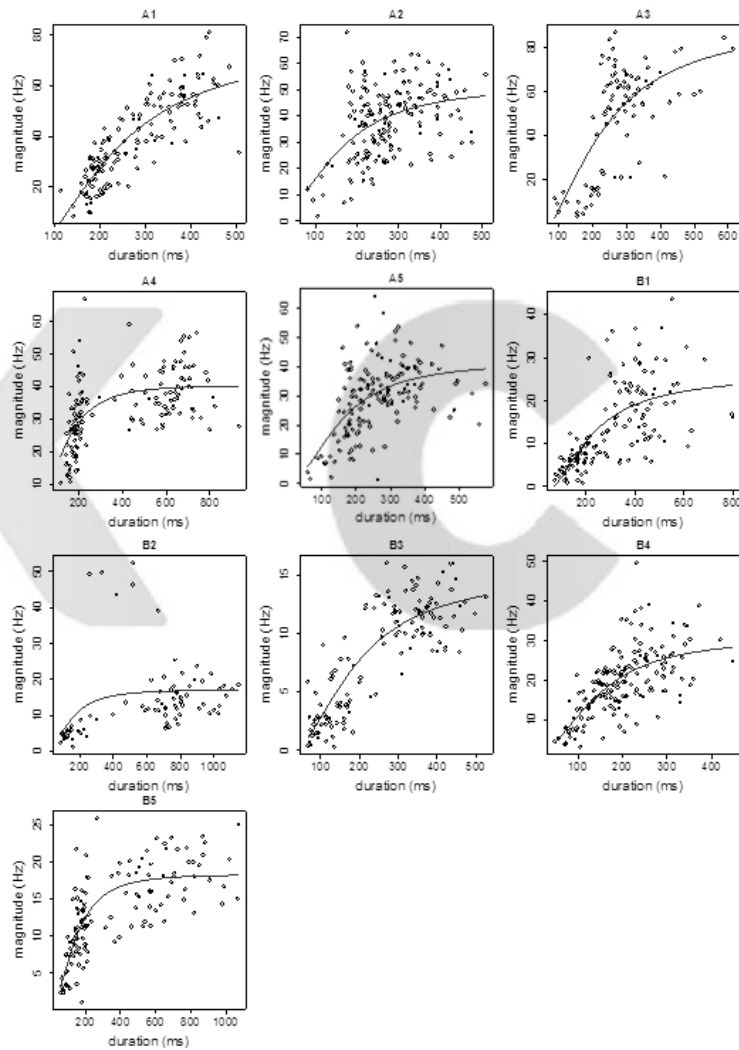

Figure 10. The relationship between magnitude and duration of the fall from H1 to L2. Curves plot the expression in (4) fitted to the data of individual speakers

Figure 10 shows the relationship between fall magnitude and fall duration for individual speakers, with curves representing the model fitted to the data. As can be seen, the model captures key properties of the data. Overall, the magnitude of the fall tends to increase as fall duration increases, and, given sufficient duration, the magnitude of the fall asymptotes to a maximum value, analyzed here as the target, T_M (e.g. A4, B2, B5) The population estimate for T_M was 42 Hz, but magnitude targets varied substantially between speaker (standard deviation of the random effect was 26 Hz). The individual estimates for T_M are closely correlated with the median of the magnitude of the rise from L1 to H1, both across and within genders. This suggests that T_M depends on the overall pitch range adopted by a speaker. The male speakers tend to have narrower pitch ranges than the female speakers but there is overlap between the groups. Slope target (T_S) was not significantly different from 0 (estimate= -0.04 Hz/ms, p=0.37). A slope target of 0 implies a preference not to change F0 and can be interpreted as a preference to minimize the effort involved in F0 production (cf. Lindblom 1963, Flemming 2001). The weight of slope target (w_S) for the population is 26997.

3.3 Plateau shortening

The analysis so far accounts for progressive undershoot of L2 as speech rate increases, culminating in realizations in which there is no dip in F0 corresponding to L2. However, we still need to account for the difference between realizations with a high plateau realization of the H1-L2-H2 sequence (Figure 5(c)), and those with a sharp peak only (Figure 5(d)). According to our analysis, extreme undershoot of L2 yields the high plateau pattern, and further time pressure results in gradual shortening of the plateau as H1 and H2 move closer together. To test this interpretation, we measured the duration of the high plateau as the interval around the F0 peak where F0 is higher than 80% of the rise from L1 to H1, as illustrated in Figure 11. According to this measure, a high plateau has a relatively long duration, while a sharp peak is shorter.
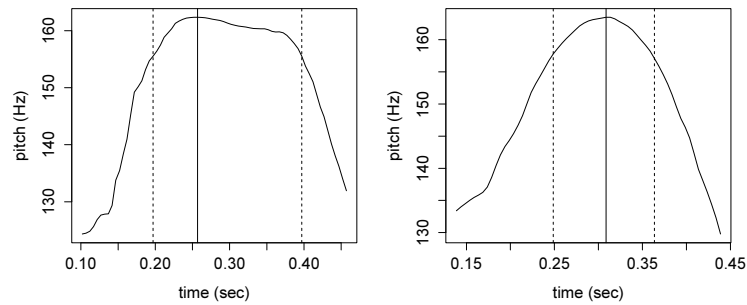
Figure 11. Measurement of plateau duration – for a plateau and for a sharp peak. The middle vertical line is the location of the F0 peak, the interval between the left and right vertical dotted lines is the measure of plateau duration.

Plateau duration is plotted against AP duration, pooled across speakers, in Figure 12, for all utterances with no F0 minimum corresponding to L2. AP duration is shorter in faster speech and longer in slower speech, so it reflects speech rate. It can be seen that plateau duration tends to shorten as AP duration decreases ($p<0.001$). We can also see that the distribution of plateau durations is continuous – that is there is not a bimodal distribution of plateau durations which could indicate a discrete distinction between presence and absence of a tone. Thus, the results are consistent with the view that the apparent deletion of L2 at high speech rates is in fact the result of compression not truncation – that is, it is the result of proximity between H1 and H2 and concomitant undershoot of L2 rather than being a result of tone deletion.
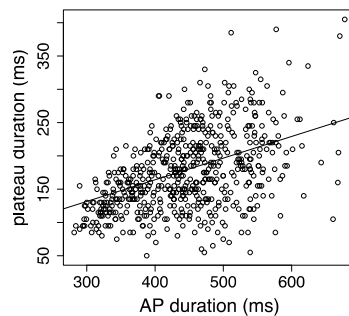

Figure 12. Plateau duration plotted against AP duration.

4. Discussion and conclusion

In this paper we have investigated whether tones can be deleted in response to time pressure due to speech rate, based on a case study of the Seoul Korean AP melody. We have seen that L2 of the L1 H1 L2 H2 Accentual Phrase melody can apparently be deleted, resulting in a rising F0 contour, but we have also seen that this is only one of several variants of the canonical melody. As speech rate increases, first we observe undershoot of the L2 target, progressing to complete undershoot of L2, yielding a high plateau, with a peaked rising F0 trajectory only occurring at the fastest speech rates.

This progression of F0 trajectory shapes can be analyzed in terms of compression processes: as speech rate increases, constraints on the timing of H1, L2 and H2 place them closer and closer together, resulting in more and more undershoot of L2. The apparent deletion of L2 is the natural end-point of these gradual compression processes, and thus does not involve tone deletion. So Seoul Korean provides no evidence that phonological tone deletion can be conditioned by speech rate.

Seoul Korean does provide a clear example of truncation which can be contrasted with the compression processes discussed so far: Tones are deleted when the AP melody is realized on APs with less than 4 syllables. Three-syllable APs are never produced with an LHLH tune, instead they are realized with only one F0 rise, i.e. a [LH] or [LHH] tune (Jun 2000, Kim 2013). The resulting F0 contours can be quite similar to the compressed pattern observed on four syllable APs at the fastest speech rate, but the compression-based analysis developed above cannot be extended to account for the pattern of truncation observed in three-syllable APs because it applies at normal and even slow speech rates and thus cannot be the result of time pressure. This pattern of truncation can be analyzed in terms of a constraint banning association of more than one tone to the same syllable, a more stringent version of the constraint operative in Hungarian (section 1)[2].

So both truncation and compression of the AP tune occur in Seoul Korean, but truncation is only conditioned by a constraint that bans more than one tone target being associated with a single syllable, whereas compression is conditioned by phonetic duration. Interestingly it is L2, which is most prone to phonetic reduction, which is deleted first in three-syllable APs, yielding an LHH tune.

This study has two further implications of general significance for the analysis of tonal phonetics and phonology: (i) The analysis shows that compression does not involve uniform shortening or reduction of the F0 contour, rather some properties of the contour are prioritized over others, resulting in greater reduction of certain tones under time pressure. In this

---

[2] It is possible to associate multiple boundary tones with an IP-final syllable (Jun 2000). This could indicate that the one-tone restriction only applies to syllables which are not subject to IP-final lengthening, or that the constraint is out-ranked by faithfulness to IP boundary tones.

case, the duration and magnitude of the initial LH rise and the alignment of L2 are prioritized over hitting the pitch target for L2 with the result that this tone can be completely obscured at fast speech rates. The relative prioritization of targets can be formalized in terms of weighted constraints. We have developed a portion of this analysis quantitatively, with reasonable fits to the experimental data. (ii) Our results show that it is not easy to distinguish compression from truncation. The tonal analysis of an F0 contour can only be determined through explicit analysis of the grammatical mapping from tones to F0 (cf. Pierrehumbert 1980).

Appendix. Speech materials

	Target AP	Carrier	
01	maɾinenin	moɾineyo	'Mary's family doesn't know'
02	naɾinenin	nollaneyo	'Nary's family is surprised'
03	mamuɾinin	miɾuneyo	'Finishing is postponed'
04	manuɾanin	moɾineyo	'A wife doesn't know'
05	monaminin	moɾineyo	'don't know Monami'
06	memoɾinin	moɾineyo	'don't know memory'
07	meɾonanin	moɾineyo	'don't know Merona'
08	nonaɾanin	mallineyo	'No–dynasty stops it'
09	lemonanin	moɾineyo	'don't know Lemona'
10	miɾinenin	moɾineyo	'don't know the milkyway'
11	miminenin	miɾuneyo	'Mimi's family postpones'
12	minaɾinin	mallineyo	'Dropwort is dried'
13	minanenin	nollaneyo	'Mina's family is surprised'
14	nunanenin	nollaneyo	'Sister's family is surprised'
15	manillanin	moɾineyo	'don't know Manila'
16	maɾimmonin	millineyo	'A lozenge is pushed'
17	noɾinnein	moɾineyo	'don't know stench'
18	amepanin	moɾineyo	'don't know ameba'
19	inajəɲin	miɾuneyo	'Lee Na–Young postpones'
20	əməninin	mallineyo	'Mother stops it'
21	oɾentʃinin	mallineyo	'Orange is dried'
22	untoŋcaɲin	moɾineyo	'don't know playground'
23	toɾirenin	millineyo	'Pulley is pushed'
24	tʃuməninin	maɾineyo	'Pocket is dried'
25	tenamunin	maɾineyo	'Bamboo is dried'
26	tallaɾanin	moɾineyo	'don't know the moon world'
27	tʃənmuntenin	millineyo	'Junior college is pushed'
28	kamnamunin	mallineyo	'Persimon tree is dried'
29	jəŋminenin	millineyo	'Youngmi's family is pushed'
30	nallaɾinin	nollaneyo	'Punk is surprised'
31	mintilenin	mallineyo	'Dandelion is dried'
32	uranjumman	moɾineyo	'don't know only uranium'

33	allatinman	miruneyo	'Only Aladin postpones'
34	tetoŋkaŋman	morɨneyo	'don't know only Taetong River'
35	mellaninman	morɨneyo	'don't know only melanin'
36	tʃiŋkɨlpelman	morɨneyo	'don't know only Jingle Bells'

REFERENCES

BOERSMA, PAUL and DAVID WEENINK. 2012. *Praat: Doing Phonetics by Computer* (Version 5.3.22) [Computer program]. retrieved 21 July 2012 from http://www.praat.org/

CASPERS, JOHANNEKE and VINCENT VAN HEUVEN. 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50, 161-71.

CHO, HYESUN. 2010. *A Weighted-constraint Model of F0 Movements*. PhD Dissertation. MIT.

_____ . 2011. The timing of phrase-initial tones in Seoul Korean: A weighted-constraint model. *Phonology* 28, 293-330.

FLEMMING, EDWARD. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18, 7-44.

FLEMMING, EDWARD and HYESUN CHO. submitted. The phonetic specification of contour tones: The rising tone in Mandarin, Ms.

GRABE, ESTHER. 1998. Pitch accent realization in English and German, *Journal of Phonetics* 26, 129-144.

GRABE, ESTHER, BRECHTJE POST, FRANCIS NOLAN and KIMBERLEY FARRAR. 2000. Pitch accent realization in four varieties of British English. *Journal of Phonetics* 28, 161-185

GRICE, MARTINE. 1995. *The Intonation of Interrogation in Palermo Italian: Implications for Intonation Theory*. Tübingen: Niemeyer.

JUN, SUN-AH. 2000. *K-ToBI (Korean ToBI) Labelling Conventions* (Version 3.1). retrived from http://www.linguistics.ucla.edu/people/jun/ktobi/k-tobi.html

KIM, KYUNGHEE. 2013. *Tone, Pitch Accent and Intonation of Korean: A Synchronic and Diachronic View*. PhD Dissertation. Universität zu Köln.

LADD, D. ROBERT. 2008. *Intonational Phonology*. Cambridge: Cambridge University Press.

LADD, D. ROBERT and ASTRID SCHEPMAN. 2003. "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics* 31, 81-112.

LINDBLOM, BJÖRN. 1963. Spectrographic study of vowel reduction. *Journal of Acoustical Society of America* 35, 1773-1781,

OHALA, J. JOHN and BRIAN W. EUKEL. 1987. Explaining the intrinsic pitch of vowels. In Robert Channon and Linda Shockey (eds.). *In Honor of Ilse Lehiste*, 207-215. Dordrecht: Foris.

PIERREHUMBERT, JANET. 1980. *The Phonology and Phonetics of English*

Intonation. PhD Dissertation. MIT.

PINHEIRO JOSÉ, BATES DOUGLAS, DEBROY SAIKAT, SARKAR DEEPAYAN and R CORE TEAM. 2015. nlme: Linear and Nonlinear Mixed Effects Models. R package (Version 3.1-121). retrived from, http://CRAN.R-project.org/package=nlme.

PRIETO, PILAR. 1998. The scaling of the L values in Spanish downstepping contours. *Journal of Phonetics* 26, 261-82.

_____. 2005. Stability effects in tonal clash contexts in Catalan. *Journal of Phonetics* 33, 215-242.

RATHCKE, TAMARA. 2013. On the neutralizing status of truncation in intonation: A perception study of boundary tones in German and Russian. *Journal of Phonetics* 41, 172-185.

R CORE TEAM. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

SILVERMAN, KIM E. A. and JANET B. PIERREHUMBERT. 1990. The timing of prenuclear high accents in English. In John Kingston and Mary Beckman (eds.). *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, 72-106. Cambridge: Cambridge University Press.

WHALEN, DOUGLAS H. and ANDREA G. LEVITT. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23, 349-366.

XU, YI and XUEJING SUN. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111.3, 1399-1413.

Hyesun Cho
Graduate School of Education
Dankook University
152, Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do
Korea 448-701
email: hscho@dankook.ac.kr

Edward Flemming
Department of Linguistics and Philosophy
Massachusetts Institute of Technology
77 Massachusetts Avenue, 32-D808
Cambridge, MA 02139
email: flemming@mit.edu