Accentual tonal targets and speaking rate in Northern Finnish

Kari Suomi
Department of Finnish,Information Studies and Logopedics,Oulu University

Abstract

This paper summarises the results of an experiment in which speakers produced the same accented target items, occurring in a constant frame sentence, at three speaking rates: Slow, Normal and Fast. The results indicate that the segmental anchoring points of the LHL tone sequence were practically the same at each speaking rate. It can be concluded that the initial L is anchored to the onset of the word-initial non-moraic consonant, the H to the end of the word's first mora, and the final L close to the end of the word's second mora. Speaking rate had little effect on the F_0 values of the tonal targets; instead,what varied was the rate of F_0 change per unit time. The results are interpreted to provide strong support for a bitonal analysis of accentuation.

Background

It has been repeatedly observed that, in Northern Finnish, for a given speaking rate and a given degree of accentuation, the form of the accentual rise-fall (or, in bitonal terms, LHL) tune is uniform cross different word structures (Suomi, Toivanen & Ylitalo, 2003; Suomi, 2005; Suomi, 2007; Suomi, submitted); earlier, we have spoken of Finnish without any qualifications, but it has now become clear that there are dialects in which the uniformity does not hold for all word structures, so it is more appropriate to talk about Northern Finnish. The rise occurs during the word's first mora, (most of) the fall during the second mora. The tonal uniformity is achieved by allophonic durational alternations in the first two syllables. For example, the four phonologically single vowels in CV.CV and CVC.CV words all have systematically different mean durations which can be characterised as [short], [long], [longish] and [very short], respectively; notice that, in each position, these vowels contrast with phonologically double vowels as these occur in e.g. CVV.CVV and CVVC.CVV words. In the studies mentioned above, no instructions were given as to speaking rate, and each talker spoke at a rate she deemed natural. The claim that the accentual tune is uniform across different word structures and that it has a certain moraic affiliation would be more convincing if it could be demonstrated that the uniformity and the moraic affiliation also hold across different speaking rates by the same speakers. This paper summarises the results of an experiment in which the same materials were spoken at three speaking rates; the detailed results will be presented elsewhere.

It was expected that, at a given speaking rate, the tones L, H and L would have the same respective target F_0 values irrespective of word structure. It was also expected that, at all speaking rates, the moraic anchoring points of the tones would be the same. As for the effect of speaking rate on F_0 values, there seemed to be at least three possibilities. Firstly, it seemed possible that the respective target F_0 values of the L, H and L tones would be constant irrespective of speaking rate, in which case the rate of F_0 changes between the target values would have to be greater in fast speech than in slow speech. Secondly, it seemed possible that the target F_0 values would be closer to each other in fast than in slow speech, the rate of F_0 change being constant across variations in speaking rate. And, thirdly, it seemed possible that speaking rate would induce some kind of a trading relationship between F_0 target values and rate of F_0 change.

The experiment

Methods

Talkers were five Northern Finnish female logopedics students. The materials were disyllabic nonsense items of the structures CV.CV, CVV.CV and CVC.CV (or $CM_1.CM_2$, $CM_1M_2.CM_3$ and $CM_1M_2.CM_3$, respectively, where M_n is the word's n^{th} mora). Consonant and vowel identities were fully counterbalanced across the materials. An example set of items is *mani, nami* (CV.CV), *maani, naami* (CVV.CV), and *manni, nammi* (CVC.CV). In addition there were items like *mina, nima; miina, niima,* etc. All of the items are phonotactically possible, and many of them are in fact real words. The orthography is fully unambiguous as to the phonemic and other structural properties of the items. There were altogether 36 different items, embedded in the constant frame sentence *Minun mielestäni ___ näyttää paremmalta* 'In my opinion ___ looks better', in which the target word was capitalised. The frame sentences and target items were to be produced at three speaking rates, resulting in 3 x 36 = 108 tokens produced by each talker.

At the beginning of the recording session, the talker received instructions, e.g. to emphasise the capitalised target item, but no instructions concerning speaking rate were given at this stage. Below, the 36 versions produced at the beginning of the recording session will be denoted as Normal speaking rate versions. After the Normal rate items had been recorded, the talker received instructions that she should next speak the same set of sentences at a 20 % faster/slower speaking rate. When this second set of the sentences had been recorded, the talker was asked to speak the same set of sentences once more, but now at a 20 % slower/faster rate than in the first round. The faster/slower rates will be denoted Fast and Slow below; one half of the talkers produced the rate versions in the order Normal-Slow-Fast, the other half in the order Normal-Fast-Slow.

Segment durations in the target items were measured, and F_0 was measured, outside the target items, at the beginning of the syllable preceding the target item and in the middle and at the end of the first syllable of the word following the target item. Within the target items, F_0 was measured at each segment boundary and in the middle of each phonetic segment; in addition, for the presumably phonetically long segments F_0 was measured half-way between the beginning and the temporal midpoint, and half-way between the temporal midpoint and the end. In addition, the temporal distance of the F_0 peak from word onset and the Hz value of the peak were measured.

Results

For one talker, there were hardly any numerical differences between the three intended Rates: 334 ms for Slow, 342 ms for Normal, and 339 ms for Fast, and the data for this speaker were discarded. For each of the remaining five talkers the numerical differences were in the intended direction, and the grand means for these speakers were 514 ms for Slow, 402 ms for Normal, and 334 ms for Fast. Thus the average Slow speaking rate was 28 % slower than the Normal rate, and the average Fast speaking rate was 17 % faster than the Normal rate.

The results are first summarised graphically in Figures 1-6 (in which the markers representing the word structures and speaking rates often hide behind each other). Figures 1-3 show that, at each speaking rate, the accentual tune was once more uniform irrespective of word structure. All Figures strongly suggest that three measurement points essentially define the tune, namely F_0 at word onset (the initial L), the F_0 peak (the H), andF_0 at word offset (the final L).

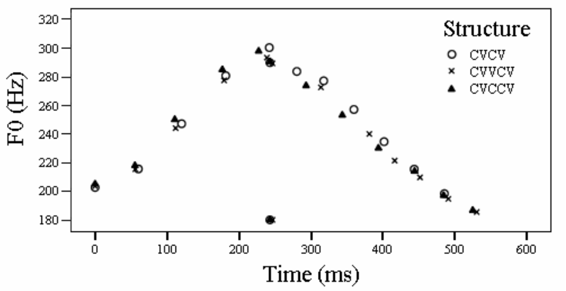

Figure 1. Mean F_0 values of the Slow Rate versions of the target items, from item onset (0 ms) to item offset. For each word structure, the end of M_1 is indicated at the bottom (arbitrarily, at the Y-axis value of 180 Hz).

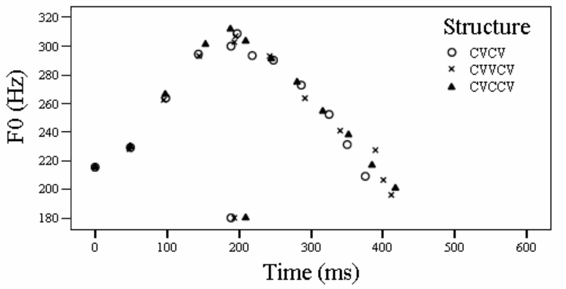

Figure 2. Mean F_0 values of the Normal Rate versions of the target items, from item onset to item offset. Further see Figure 1.

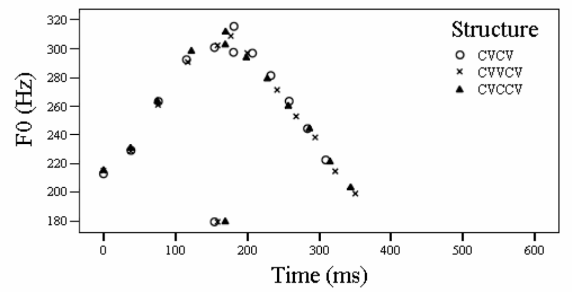

Figure 3. Mean F_0 values of the Fast Rate versions of the target items, from item onset to item offset. Further see Figure 1.

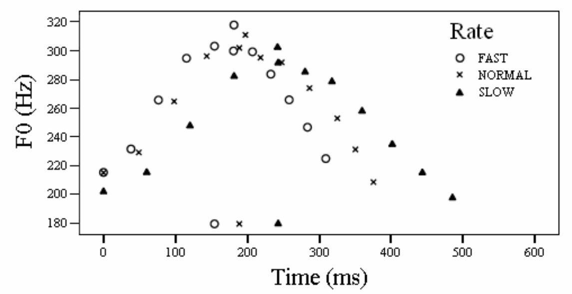

Figure 4. Mean F_0 values of the CV.CV items in the three speaking rates. Further see Figure 1.

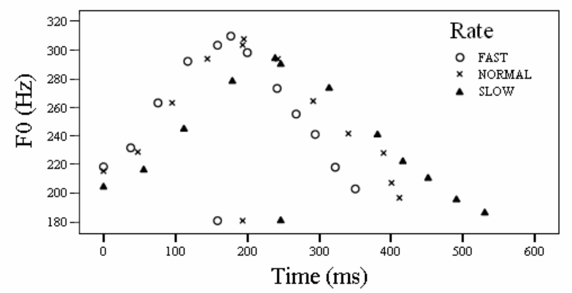

Figure 5. Mean F_0 values of the CVV.CV items in the three speaking rates. Further see Figure 1.

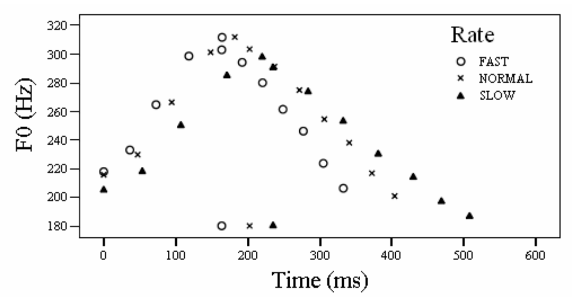

Figure 6. Mean F_0 values of the CVC.CV items in the three speaking rates. Further see Figure 1.

Neither word structure nor speaking rate had any effect on the Hz values at word onset or at the peak, nor on the mean differences between the temporal distance of the F_0 peak from word onset and the distance from word onset to the end of M_1. Thus the peak F_0 always coincided with the end of M_1 irrespective of structure and speaking rate. Moreover, there were no effects on the Hz difference between word onset and the peak. Up to the F_0 peak, then, there were no significant effects on the Hz values.

On the Hz values at word offset, speaking rate had a main effect as did, very narrowly, word structure, and there was no interaction. Post hoc tests indicated that the mean Hz value was lower in the Slow versions than in the Fast ones, and that the Normal versions did not differ from the other versions. For Structure the post hoc tests indicated no differences among the three structures, despite the narrowly significant main effect. The means were 211 Hz, 195 Hz and 198 Hz for the CV.CV, CVV.CV and CVC.CV items, respectively, and presumably the statistical outcome can be interpreted to indicate that there was a strong tendency for F_0 to be higher at item offset in the shorter CV.CV items than in the others. On the other hand, there were no effects either on the F_0 difference between the peak and word offset, or on the F_0 difference between the peak and the middle of the syllable following the target items. At the middle of the post-target syllable, the F_0 means ranged from 182 Hz to 183 Hz as a function of structure, and from 179 Hz to 185 Hz as a function of Rate. Thus, it seems that any differences in Hz values at word offset were levelled at the middle of the following syllable.

Discussion

The results show that, at a given speaking rate, the accentual tune is uniform irrespective of word structure, as has been consistently observed earlier. In a more novel vein, the results also show that the accentual tune is realised in essentially the same with respect to its anchoring points irrespective of speaking rate; of course, the tune is not uniform for different speaking rates, but varies in its durational properties as a function of rate-dependent segment durations. Speaking rate did not affect F_0 values except at word offset, and here it did so only marginally, and at the middle of the following syllable, the rate effect was no longer effective. Thus, the tonal F_0 targets seem to be invariant irrespective of speaking rate, and what varies is the rate of F_0 change per unit time. This circumstance strongly suggests that the F_0 curves are determined by invariant tonal targets. We interpret this as strong support for the bitonal analysis of accentuation as proposed by Bruce (1977), and later by others.

Irrespective of speaking rate, then, the segmental anchoring points of the LHL tone sequence remained practically the same. The initial L is unquestionably anchored to the onset of the word-initial, non-moraic consonant, and the anchoring point of the H is the end of M_1. The precise anchoring point of the final L remains somewhat unclear. A difficulty here is to properly distinguish between accentual and intonational effects on F_0. In Suomi (2007), in which the target words occurred in the middle of rather long utterances, an F_0 plateau was reached by the middle of the third syllable of tri- and tetrasyllabic words, and in short words the accentual fall continued in the next, unaccented word. In the present experiment, in which the target items occurred in a rather short utterance, F_0 continued to gradually fall after word offset, without reaching a plateau, presumably because of falling sentence intonation. These differences notwithstanding, most of the fall from the H to the L occurs, in all experiments reviewed here, during M_2 in all word structures. This question requires further clarification, but it seems legitimate to conclude that speaking rate has no radical effect on the anchoring point of the final L, which is close to the end of M_2. In summary, it can be concluded that speaking rate has little effect on the segmental anchoring points of the accentual LHL tune in Northern Finnish.

In Northern Finnish, then, given a constant speaking rate, the accentual tune has a uniform shape irrespective of word structure. When speaking rate is varied, segment durations and the total duration of the accentual tune naturally also vary, but the segmental anchoring points remain the same (with the reservation that the exact anchoring point of the final L is not known). The tonal uniformity, at a given speaking rate, is achieved by subphonemic durational alternations. This is in contrast to many other languages in which the tonal realisation of accent varies as a function of the segmental composition of the accented syllable. Thus e.g. Arvaniti, Ladd & Mennen (1998) and Ladd, Faulkner, Faulkner & Schepman (1999) report, for Greek and Standard Southern British English, respectively, that the slope and duration of the (prenuclear) accentual tonal movement vary as a function of the structure of the accented syllable. It seems paradoxical that, in a full-fledged quantity language like Finnish, in which segment durations signal lexically important distinctions (as well as inflectional distinctions), segment durations nevertheless also vary to serve tonal purposes, while in non-quantity languages the segmental composition of the accented syllable in contrast determines the tonal realisation.

But perhaps, as is argued in more detail in Suomi (submitted), there is no paradox. Perhaps the moraic alignment of accent and its uniformity across different word structures, although these circumstances entail durational alternations, yet bring order to what might otherwise be a chaos, by providing a frame of reference against which it is easier to maintain the important quantity distinctions than would be in a system in which the segmental anchoring of accent was much looser? For example, a rise-fall tune during the initial syllable and a low continuation of the fall during the second syllable strongly suggest that the initial syllable is heavy (as in eg. *tuuli* ‘wind’ and *tulli* ‘customs’), whereas a rise during the first syllable followed by a deep fall from high to low during the second syllable strongly suggest that the initial syllable is light (as in *tuli* ‘fire’). Perhaps, if the shape of the accentual tune were allowed to vary as a function of the structure of the stressed syllable of the accented word, it would be too difficult, without support from tonal cues, to maintain the quantity opposition in both vowel and consonants?

References

Arvaniti A, Ladd D R, Mennen I (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetic*s, 26: 3-25.

Bruce G (1977). *Swedish word accents in a sentence perspective*. Sweden: Gleerup.

Ladd D R, Faulkner D, Faulkner H, Mennen I (1999). Constant ”segmental anchoring” of F_0 movements under changes in speech rate. *Journal of the Acoustical Society of America,* 106: 1543-1554.

Suomi K, Toivanen J, Ylitalo R (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, 31: 113-138.

Suomi K (2005). Temporal conspiracies for a tonal end: Segmental durations and accentual f_0 movement in a quantity language. *Journal of Phonetics*, 33: 291-309.

Suomi K (2007). On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics*, 35: 40-55.

Suomi K (submitted). Prosodic interactions in a quantity language: Accent, quantity and segment durations in Northern Finnish.