QUANTITY AND TONE IN FINNISH LEXICALLY STRESSED SYLLABLES

Martti Vainio[1], *Daniel Aalto*[2], *Juhani Järvikivi*[3], *and Antti Suni*[1]

[1]Department of Speech Sciences
University of Helsinki, Helsinki, Finland
[2]Institute of Mathematics
Helsinki University of Technology, Finland
[3]Department of Psychology
University of Turku, Finland
martti.vainio@helsinki.fi

ABSTRACT

This paper presents results from a study on the tonal aspects of quantity in Finnish lexically stressed syllables. Fourteen speakers produced a set of 66 utterances where the quantity and structure of the lexically stressed syllable was systematically varied. The tonal aspects of the syllable nucleus and nucleus and coda in case of closed syllables was studied in the framework of the Target Approximation theory as formulated by Yi Xu. The results show a clear tendency towards the quantity distinction and bimoracity in general in Finnish to be signalled tonally by a dynamic falling tone as opposed to a static high tone in short (one mora) nuclei.

1. INTRODUCTION

Finnish is a full-fledged quantity language with two degrees of length for practically all sounds in the language and applicable in almost all positions within a word. The basic word structure in Finnish consists of a disyllabic sequence of a stressed and unstressed syllables which, furthermore, form the basic rhythmic unit; the metric foot. The lexical stress on the first syllable of the word is usually signaled by longer segmental durations and higher f_0; either rising (short vowel in the nucleus) or rising-falling (long vowel in the nucleus). In a sense, there is a tonal contrast between the two syllables with the unstressed syllable usually being lower in f_0 than the stressed one. In addition to signaling word structure, the local pitch changes serve as strong cues to prominence in Finnish [1] and a rise within the stressed syllable has been linked to the perception of word boundaries [2].

It is a well established fact that in Finnish the f_0 curve peaks occur during a long stressed vowel but tend to occur at the end of the short vowel or immediately after it [3, 4, 5]. It has been argued that the prosodic structure of the words is the same and that the longer duration of the long vowel allows for the peak to occur relatively earlier [4]. The earlier peak has been shown, however, to be relevant for the perception of quantity in both Finnish [5] and Estonian [6]. That is, the peak location is not related to the word alone, but is an important cue to quantity. It is, therefore, important to see whether these tonal differences are due to similar articulations or if there are, in fact, different underlying structures that are being signalled tonally during the stressed syllable.

Yi Xu (see [7] for a review) has formulated a theory for speech melody based on syllable synchronic pitch targets. According to his theory the bisyllabic sequence of Finnish could be analyzed as a sequence of two tonal targets. Since the second (unstressed) syllable is almost always lower than the first one, it can be analyzed as a low tone. Moreover, it always approximates a static target as the f_0 movement at the end of the syllable is usually horizontal. The tonal differences between the long and short (two vs. one moraic, respectively) stressed syllables could be explained by their having two distinct tonal targets. According to Xu, the tonal shape of the peak and its surroundings could be formed by either a rise toward a static high target or a falling tone, which causes a rise (peak) and a falling pattern during the syllable. The underlying targets are best seen at the end of the syllable when the f_0 approximates the target and by the shape of the f_0 contour during the syllable.

Thus, it could be that even in a quantity language like Finnish long stressed vowels (and possibly bimoraic syllables in general) are being signalled by a dynamic falling tone as opposed to static high tone in short (one mora) stressed syllable. More generally, the tonal structure of words in Finnish could be characterized by a tonal opposition within a disyllabic, or possibly a tri-syllabic (a trochaic foot in metrical terms), sequence of H/F versus L and a dynamic vs. static tonal opposition within the stressed syllable of the word.

In order to assess the hypothesis that quantity is signalled by tonal means we ran a production experiment using two types of target word pairs (CV.CV vs. CVV.CV and CV.CV vs. CVC.CV) embedded in carrier sentences. The sentences with the target words were then read aloud, recorded and analyzed with regard to tonal structure of the stressed syllables of the target words. To analyze the target words tonally we calculated the derivatives of the f_0 curve during the stressed syllable and compared the results between the different syllable structures.

2. MATERIALS AND PROCEDURE

Four sets of words consisting of nouns and adjectives were selected from the Turun Sanomat lexical database [8] consisting of 22.7 million words as follows: Twenty-four word pairs such as /pu.ro/ - /puu.ro/ ("stream"- "porridge") were selected for the CV1-CVV conditions, and twenty-one pairs were selected for the CV2-CVC condition, e.g., /ka.ma/ - /kam.pa/ ("stuff" - "comb").

In the first condition the words differed only for the length of the vowel in the first syllable. In the second condition the words in each pair differed only in that the sound starting the second syllable of the first word and ending the first syllable of the second word in each pair was always the same ([l],[r],[m],[n]) as in the example above. Using liquids and nasals at the end of the bimoraic syllable ensured that the f_0 could be detected throughout the syllable and only in the case of the trill [r] were there any difficulties.

In order to avoid any potential effects caused by differences in familiarity, the words in the two sets were matched for lemma frequency (CV1 and CVV: mean frequency, 34 and 26 per million, respectively $[t - test, t < 1]$; CV2 and CVC: mean frequency, 40 and 42 per million, respectively). Each word pair was embedded in a sentence frame so that the surrounding context was kept either identical or as similar as possible. In all sentence contexts at least the words that immediately preceded and followed the target were always identical. Altogether 45 sentence frames were created, a different sentence frame for each experimental pair. The syntactic positions of the target words in the sentence frames as well as the length of the sentence frames were also controlled for.

The sentences containing the target words were counterbalanced between two experimental lists with each list containing only one of the target word per word pair. Both lists had an equal number of target sentences from each condition. In addition, 21 filler sentences were added to both lists.

Fourteen native speakers of Finnish (13 female) were assigned to one of the lists of 66 sentences (7 per list) in the order of appearance. The speakers were all students at the University of Helsinki and none reported any hearing problems. Only one of the speakers was familiar with speech prosody research. The speakers were all paid for their participation.

The speakers were asked to read the sentences aloud one by one in a normal speech rate. The recording was done in a sound proof room at the Department of Speech Sciences using a high quality condenser microphone placed approximately 20 cm from the subjects mouth. The sound was stored on a computer hard drive using a high quality AD converter. The recordings were then split into utterance sized chunks and manually segmented and labelled. For this study either the nucleus of the stressed syllable was marked (CV and CVV cases) or the nucleus and the coda (CVC or closed syllable cases). The raw pitch values were then extracted for further analysis.

3. DATA ANALYSIS

The extracted pitch values were loaded in Matlab 7.1.0.183 (R14) and converted in semitones. To estimate the underlying pitch targets the slope of the f_0 curve was calculated at the end of each marked segment. To avoid too heavy dependence on any particular f_0 value the derivative was smoothened by taking averages. More precisely, the end point of a segment was connected with seven preceding points in the f_0 curve and the average of the slopes of the resulting seven secants. This corresponds to a time window of 38.5 ms with standard deviation of 20 ms. Since the mean duration of the first segment (nucleus) was 120 ms with standard deviation of 5 ms, the calculated target approximation reflects the behaviour of the fundamental frequency during the last quarter of the segment. Lastly, the derivative was scaled with the segment duration to give the target change rate in semitones per segment.

The situation is illustrated in the Figures 1 and 2. The solid lines, the dots and the vertical dotted lines represent approximate targets, pitch with respect to speakers mean f_0 and segment boundaries respectively.

4. RESULTS

The results from the production experiment are depicted if Figure 3. With respect to the target change rate as the dependent measure, 2 x 2 analyses of variance were carried out for the subject (F1) and item (F2) means using Syllable Length (CV vs. CVV/CVC) and Type of Syllable Contrast (cv1-cvv vs. cv2-cvc) as within-subject factors. In the item analyses Syllable Length was treated as a within-item and Type of Syllable Contrast a between-item factor.

The results showed a significant main effect of syllable length both by subjects and by items $[F1(1, 13) = 56.50, p < .001; F2(1, 43) = 94.41, p < .001]$. However, there was no effect of syllable type $[F1 \text{and} F2 < 1]$ and no significant interaction $[Fs < 1]$, suggesting that only the syllable length

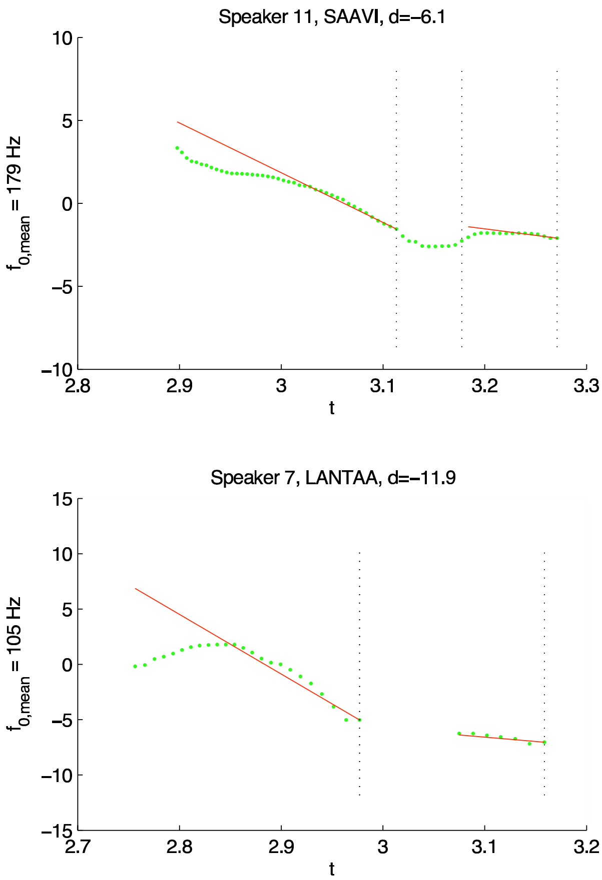

Fig. 1. Typical falling f_0 patterns for bimoraic bisyllabic words. The vertical lines depict boundaries between the end of the first syllable (first line) and the end of the onset of the second syllable (second line) and the end of the nucleus of the second syllable (third line). The vertical lines also show the f_0 range of the whole utterance. The grey dots depict the one speakers mean f_0 curve and the straight lines stand for the estimated tonal targets.

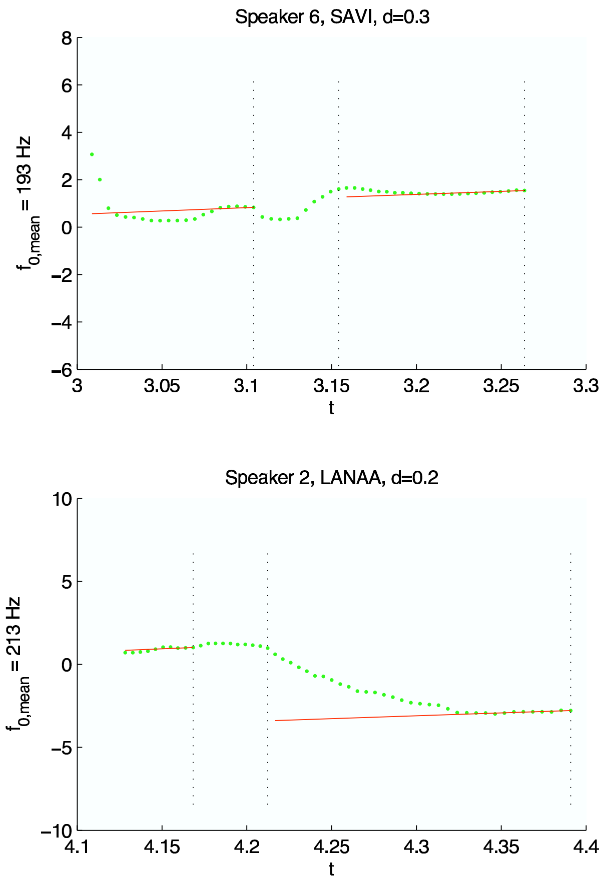

Fig. 2. Typical static High vs. Low f_0 patterns for bisyllabic words with one mora in the stressed syllable. The higher second syllable in the upper pane is probably a boundary tone before a relative clause signaling continuity.

was a decisive factor in the experiment. Moreover, the lack of any trace of interaction shows that both CVV and CVC behaved identically in terms of tonality, despite the segmental phonological difference (as also shown in Figure 3.).

5. DISCUSSION

The estimation of the target by averaged derivatives gives more emphasis on the final point of the segment compared to linear curve fitting. This obviously leaves any segmental influence to the f_0 curve in the results. For instance the fact that the average derivatives of the CV syllables are slightly negative can be, tentatively, explained by decreasing airflow in the glottis as the upper vocal tract is narrowing towards a closure.

6. CONCLUSION

The general and tentative hypothesis for Finnish tonal structure of words is then as follows: there is a tonal opposition within a disyllabic, or possibly a tri-syllabic (a trochaic foot in metrical terms), sequence of H/F versus L and a dynamic vs. static tonal opposition within the stressed syllable of the word. The nature of the unstressed syllables within a metrical feet is still somewhat uncertain and needs to be studied further. Impressionistic observations seem to suggest that there are further differences between the unstressed syllables of the trochaic foot. Moreover, the tonal opposition between a static and dynamic tones may not be limited to stressed syllables only.

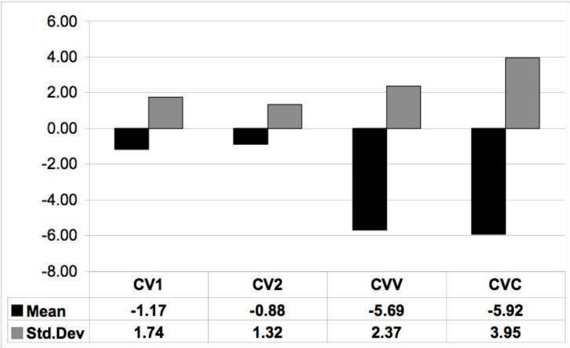

Fig. 3. Summary of the results for the different syllable structures. Both means and standard deviations are shown.

It should also be noted that none of the examples in Figures 1 and 2 show the typical pointed hat pattern considered typical for Finnish accentuation. This is in contrast with Suomi [4], who argues that there is a tonal uniformity underlying different word and syllable structures. The obvious un-uniformity in our results may be due to the fact that we did not control for accentuation in the material and most target words were only moderately accented. It is probable that, had we used narrowly focused words only, our results would have resembled those of Suomi's.

Based on the present study alone, it cannot be determined whether the tonal cues are by themselves sufficient to signal the quantity opposition. However, the results do show that tonality is an integral part of the phonetic structure of Finnish words and syllables. They further suggest that tonality may even have a function akin to phonological in Finnish. Whether this is an emerging tonal pattern or one that quantity has always built on remains to be studied cross-linguistically.

7. ACKNOWLEDGEMENTS

We would like to thank Leena Wahlberg for her help in recording and preparing the data. The present study was supported by grant no. 107606 from the Academy of Finland to the first author and grant no. 106418 from the Academy of Finland to the third author.

8. REFERENCES

[1] Martti Vainio and Juhani Järvikivi. Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics*. In Press.

[2] Jyrki Tuomainen. *Language specific cues to segmentation of spoken words in Finnish: Behavioral and event-related brain potential studies*. PhD thesis, de Katholieke Universiteit Brabant, 2001.

[3] Jaakko Lehtonen. *Aspects of Quantity in Standard Finnish*. Number VI in Studia Philologica Jyväskyläensia, 1970.

[4] Kari Suomi. Temporal conspiracies for a tonal end: Segmental durations and accentual f0 movement in a quantity language. *Journal of Phonetics*, 33:291–309, 2005.

[5] Michael O'Dell. *Intrinsic Timing and Quantity in Finnish*. PhD thesis, University of Tampere, 2003.

[6] Ilse Lehiste. Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 4:113–117, 1976.

[7] Yi Xu. Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46:220–251, 2005.

[8] M. Laine and P. Virtanen. *Turun Sanomat computerised lexical data base (database and program)*. Unpublished, 1996.