PERCEPTION OF VOWEL LENGTH: TONALITY CUES CATEGORIZATION EVEN IN A QUANTITY LANGUAGE

Juhani Järvikivi[1], Daniel Aalto[2], Reijo Aulanko[3] & Martti Vainio[3]

University of Turku[1], Helsinki University of Technology[2], University of Helsinki[3]
juhani.jarvikivi@utu.fi, daniel.aalto@hut.fi, reijo.aulanko@helsinki.fi, martti.vainio@helsinki.fi

ABSTRACT

A two-alternative forced-choice categorization experiment (2AFC) tested whether the type of tone (static high vs. dynamic fall) affected the perception of the length of a stressed initial syllable in Finnish, when the participants had to categorize it as "short" or "long". In addition to the main effects of the duration of the first and second syllables, the results showed a significant main effect of tone that was qualified by an interaction with the duration of the first syllable nuclei. More precisely, the participants were *ceteris paribus* more likely to categorize the vowel of the first syllable as "long" in the dynamic fall condition than the high tone condition. The results showed that, alongside with duration, also the tonal structure is used as a strong perceptual cue for the quantity opposition in Finnish.

Keywords: Quantity, Duration, Perception, Tone, Finnish.

1. INTRODUCTION

Finnish is a prototypical quantity language that exhibits two degrees of length for most sounds of the language in most positions. Thus, virtually any phoneme can be realized in the short and the long form. Examples abound: tule – tuule – (ei) tulle – (ei) tuulle – tulee – tuulee – tullee – tuullee. The lexical stress always falls on the first syllable of the word and is usually signaled by longer segmental durations and higher f0; either rising (short vowel in the nucleus) or rising-falling (long vowel in the nucleus). It has been shown that in Finnish the f0 curve peaks occur during a long stressed vowel but tend to occur at the end of the short vowel or immediately after it. The standard argument has been that the overall 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 [1, 2, 3].

In average the duration of a long phoneme is roughly double the length of the short one. Nevertheless the durational distributions of the two quantity categories overlap. According to Lehtonen [1] the quantity difference can be explained by the durations of different segments and their mutual relationships. In his view no tonal differences exist between the vowels of the two quantity categories.

This is in contrast with the view of Malmberg [4], that the tonal shape differs according to the category. He based his argument on one reader and a couple of sentences. The target words were in non-stressed positions and the f0 trajectories were extracted manually from the wave form given by a kymograph. Based on this, Malmberg [4] concludes that the long vowels are characterized by a descending pitch in the second half of the vowel and an ascending pitch in the first half in a few cases, but an even one in most of the other ones. The short vowels, in turn, have a rising pitch in the beginning, and then their pitch is essentially united. Recently, Suomi [2] presented evidence from Finnish bisyllabic words. He argued that there exists a uniform tonal pattern which is timed according to morae: in the first mora of a word there is a rise in pitch, whereas in the second there is a fall. Thus, taken this way the results of Suomi suggest that underlyingly all syllable and word types are uniform with respect to their tonal form [2].

Vainio et al. [5] studied the production of the quantity opposition in Finnish using CV-CVV and CV-CVC word pairs, such as /pu.ro/ - /puu.ro/ ("stream"- "porridge") and /ka.ma/ - /kam.pa/ ("stuff" - "comb"). Their results showed a clear tendency towards the quantity distinction (and bimoracity in general) to be signaled tonally by a dynamic falling tone as opposed to a static high tone in short (one mora) nuclei. To explain their results, Vainio et al. [5] propose a model based on the target approximation framework by Xu (see [6] for an overview). Thus, the typical tonal pattern of a bisyllabic word is taken to consist of two consecutive targets: high (H)/fall (F) and low (L). The hypothesis is that the tonal target varies

according to the length of the syllable. Thus, an open syllable with a short or long vowel would get the target H or F, respectively. As a result, both the observations of Malmberg [4] and Suomi [2] may be subsumed as special cases under the proposed target approximation framework: the rise-fall pattern is a typical realization of the falling tonal target when the initial pitch height is relatively low. This happens when the word is in a stressed position, like in Suomi's study [2]. When the stress is weak, there is no rise, but the syllable's final fall remains. In Vainio et al. [5] the tonal characterization of Finnish was supported by statistical analyses of a large number of words in production. It is of interest, however, whether the observations of Vainio et al. generalize to perception. Previous evidence from Estonian [7] and Finnish [3] suggests that this may indeed be the case. As the evidence from Finnish is only suggestive, we set out to investigate this issue in a two-alternative forced-choice categorization experiment, where the participants had to judge the length of the vowel in the first syllable of a two-syllable word as either "short" or "long". We hypothesized that if tonality is used to cue the perception of the degree of length in addition to duration in Finnish, we should observe a clear tendency of the dynamic falling tone to induce more categorizations in the "long" than in the "short" category as compared to the static high tone, when the syllable duration is controlled for.

2. EXPERIMENT

2.1. Method

2.1.1. Participants

Nineteen students from the University of Helsinki participated in the experiment. None reported any hearing problems.

2.1.2. Materials

The materials for the experiment consisted of five pairs of words which differed in the quantity of the first syllable vowels; short (CV) or long (CVV). Orthographically the pairs were as follows: sika/siika (pig/perch), kisu/kiisu (kitten/ore), Mika/Miika (Finnish male forenames), kato/kaato (loss/felling), and pika/piika (instant/maiden). The ten words were spoken within a carrier sentence ("sano sana X tasaisesti"; 'say word X evenly') by a female speaker. The utterances were recorded using a high quality condenser microphone (AKG 4000B) and a high quality analogue to digital converter (Digidesing Digi002). The word segments were then isolated from the carrier sentences for further processing which was done by the PSOLA algorithm in the Praat program [8].

The syllable nuclei of the words were first labeled for both f0 and duration manipulation. The duration of each first syllable nucleus was set to five different values with a 25 millisecond difference between the values (75, 100, 125, 150, and 175 ms). The actual durations of the vowels were, however, somewhat longer (15-20 ms) due to the fact that the manipulations did not take place throughout the whole vowel. The duration of the second syllable vowel was set at 50 and 100 ms (again with somewhat longer actual durations remaining in the stimuli). The tone of the first syllable vocalic nucleus was set at two different levels: a straight linear (high) tone throughout the vowel or a linear fall. Figures 1 and 2 show two examples of the resulting stimuli using the word "pika" as the source of manipulation. The f0 of the second syllable was not changed. All in all, this resulted in 10 X 5 X 2 X 2 = 200 stimuli. This still left the second syllable onset consonant intact. The mean durations of the second syllable onset consonants were 146 ms for the CV words and 120 ms for the CVV words.

Figure 1: Word "pika" with a high, static tone on the first syllable and 100 ms duration on both syllabic nuclei.

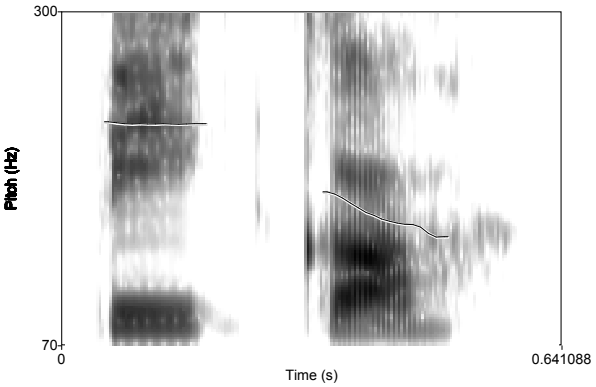
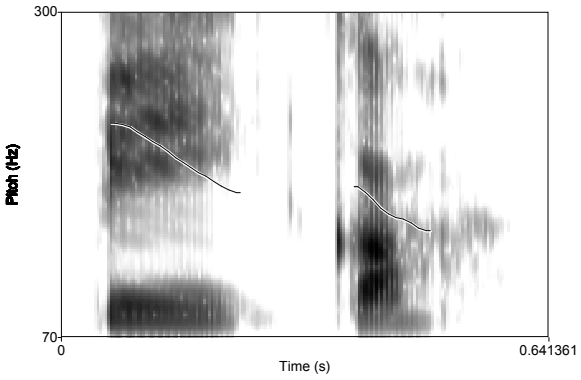


Figure 2: Word "pika" [>/piika/] with a falling dynamic tone on the first syllable and 150 ms duration on the first and 50 ms duration on the second syllabic nuclei.



2.1.3. Procedure

The resulting 200 stimuli were collected together into a long sound file in a randomized order. A two-second silence was inserted between each stimulus (resulting in approximately 2.2 s interval between the trials). After every ten stimuli, a short tone was played in order for the participants to be able to know whether they were in synchrony with the test. The participants were given a sheet of paper with numbered lines containing the stimulus number and two boxes. They were instructed to mark on the sheet whether they perceived the vowel of the first syllable of the stimulus word to be either "short" or "long". A practice block consisting of ten trials preceded the experimental trials.

2.2. Results and discussion

In what follows we will report the results only for the half of the materials that were constructed using the words with a short (CV) first syllable as the basis. This is because in these materials the segmental information, especially the duration of the initial consonant in the second syllable, is most in line with "short" rather than "long" decision, and thus offers the strongest test case for the influence of tonal structure.

The data was analyzed using the proportion of "long" responses as the dependent measure. The data was then subjected to 2 X 2 X 5 Analysis of Variance (ANOVA) with the factors Tone (high vs. fall), Duration of the Second Syllable (50 vs. 100 ms), and Duration of the First Syllable (75, 100, 125, 150, 175 ms) as within-participant variables. The results from the experiment are depicted in Figures 3 (type of tone x 1st syllable vowel duration) and 4 (2nd syllable vowel duration x 1st syllable vowel duration).

Figure 3: The proportion of "long" responses (Y-axis) as a function of the first syllable vowel duration for the stimuli with the dynamic falling tone (intact line) and the high tone (dotted line).

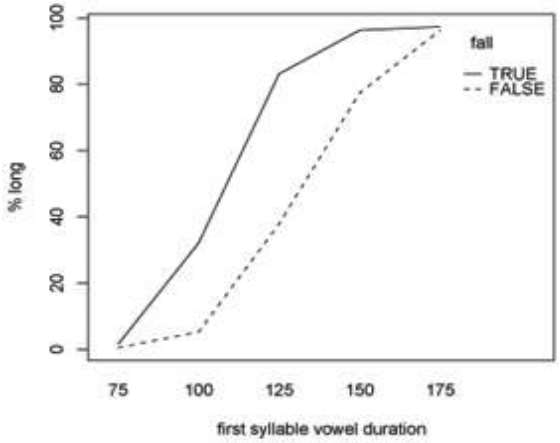
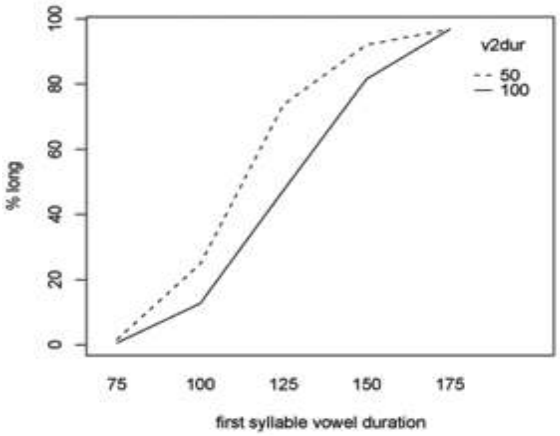


Figure 4: The proportion of "long" responses (Y-axis) as a function of the first syllable vowel duration for the stimuli with the long 100 ms second syllable vowel duration (intact line) and the short 50 ms second syllable vowel duration (dotted line).



The ANOVAs revealed significant main effects for all three factors (F- and p-values and effect sizes in partial *Eta* squared): Tone [$F(1,18) = 139.93$, $p < .0001$, $\eta_p^2 = .886$]; Duration of the Second Syllable [$F(1,18) = 72.46$, $p < .0001$, $\eta_p^2 = .801$]; Duration of the First Syllable [$F(4,72) = 375.17$, $p < .0001$, $\eta_p^2 = .954$].

Thus, as expected, the results showed that the perception of the first syllable as long dependent most strongly on its duration: the longer the duration the more likely it was perceived as long. However, the results clearly showed that also the type of tone played an important role: the first syllable was significantly more likely to be perceived as long with the falling tone than the high tone. In contrast, the duration of the second syllable had an inverse effect as compared with the duration of the first syllable: the shorter 50 ms duration induced significantly more "long" responses than the longer 100 ms duration. However, this main effect was also qualified by a significant interaction between the durations of the first and second syllables [$F(4, 72) = 15.31$, $p < .001$, $\eta_p^2 = .460$] showing that the duration of the second syllable had a significant effect only when the duration of the first syllable was 100-150 ms (all ts > 3.70, ps < .005) but not in the two extremes (75 and 175 ms: ts < 1.50, ps > .16).

The effect of the type of tone (fall vs. high) was not modulated by the duration of the second syllable ($F < 1$). However, the type of tone did interact with the duration of the first syllable [$F(4,72) = 25.140$, $p < .001$, $\eta_p^2 = .585$]. Analyses of simple effects showed that the type of tone did not affect the perception of the length of the first syllable in the two extreme categories (75 and 175 ms: ts < 1), but did have a significant effect in the three middle categories (100 ms – $t(18) = 5.12$, $p < .001$; 125 ms – $t(18) = 9.56$, $p < .0001$; 150ms – $t(18) = 5.07$, $p < .001$). As can also be seen in Figure 3, the falling tone acted as a strong cue towards perceiving the first syllable as long rather than short when the duration of the syllable itself did not offer unambiguous enough support for a clear "short" or "long" categorization. Quite naturally, the effect of the type of tone was also clearest with the most ambivalent 125 ms duration of the first syllable. We also found an additional significant 3-way interaction [$F(4,72) = 3.29$, $p < .05$, $\eta_p^2 = .155$] that was most likely the result of the duration of the second syllable interacting with the duration of the first syllable but not affecting the influence of the of tone.

In order to further assess the effects of the three variables, we ran logistic regression analyses with the duration of the first syllable, duration of the second syllable and tone as independent variables and the proportion of "long" decisions as the dependent variable.

The results from the logistic regression analysis clearly support the ANOVA results; i.e., the most important predictor for quantity being the duration of the first syllable vowel [$\chi^2(16) = 76.49$, $p < .0001$] followed by tone [$\chi^2(10) = 49.77$, $p < .0001$] and second syllable vowel duration [$\chi^2(10) = 26.16$, $p < .005$]. As expected, the first syllable vowel duration and tone interaction was also significant [$\chi^2(8) = 21.64$, $p < .01$]. This time, however, no other interactions reached statistical significance. The whole model was highly significant [$\chi^2(19) = 77.75$, $p < .0001$, $R^2(19) = .909$] and explains more than 90% of the variance in the responses.

3. CONCLUSION

The results support the view that the quantity distinction in a stressed syllable in Finnish is systematically signaled with tonal means in addition to and in interaction with durational means. Taken together with the compensatory shortening of the following unstressed syllable, the falling tone allows for the bisyllabic sequence - and subsequently the whole word - to be shorter in duration than it would be without the tone. Thus, it is possible that a further function of the tone, in addition to providing means for accentuation and increased prominence, is to compensate for the increased duration, thus saving time.

4. REFERENCES

[1] Lehtonen, J. 1970. *Aspects of quantity in standard Finnish*. Studia Philological Jyväskyläensia 6. Jyväskylä: University of Jyväskylä.

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

[3] O'Dell, M. 2003. *Intrinsic timing and quantity in Finnish*. Acta Universitatis Tamperensis 979. Tampere: Tampere University Press.

[4] Malmberg, B. 1949. Review of "Voyelles longues et voyelles brèves." by Marguerite Durand. *Studia Linguistica*, 3(1), 39–59

[5] Vainio, M., Aalto, D., Järvikivi, J., Suni, A. 2006. Quantity and tone in Finnish lexically stressed syllables. In: A. Belotel-Grenié & M. Grenié, (eds.), *Proceedings of the 2nd International Symposium on Tonal Aspects of Languages*, 121-124.

[6] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46, 220-25.

[7] Lehiste, I. 2004. Bisyllabicity and tone. In: *International Symposium on Tonal Aspects of Languages*, *With Emphasis on Tone Languages*, Beijing, China, March 28-31, 2004, 111-114.

[8] Boersma, P. 2001. PRAAT, a system for doing phonetics by computer. Glot International, 10, 341–345.