# EFFECT OF CONTEXT *F*0 ON QUANTITY IN FINNISH

Michael O'Dell[1], Tommi Nieminen[1,2], Joonas Vakkilainen[2]

[1]University of Helsinki, [2]Tampere University
michael.odell@helsinki.fi, tommi.k.nieminen@tuni.fi, joonas.vakkilainen@tuni.fi

## ABSTRACT

There is evidence that perceived quantity in Finnish is affected by $F0$ manipulation, at least within stressed vowels in accented position. We investigated whether unstressed vowels would be affected by $F0$ manipulation outside the vowel. Recordings of two Finnish sentences were manipulated for total duration and $F0$ range of context, keeping unstressed target vowel duration and $F0$ constant. Native subjects then judged whether the target vowel was long or short. Results show that a long context sentence makes a constant target vowel sound short and vice versa, but more importantly, that changes in $F0$ range can also shift the quantity boundary of an unstressed vowel, even though the $F0$ contour of the vowel itself does not change. This suggests that $F0$ can interact with quantity because manipulations which change perceived tempo will also affect quantity perception, which is judged relative to tempo rather than being based on absolute durations.

**Keywords:** Finnish, quantity, $F0$, tempo

## 1. INTRODUCTION

Finnish is a full fledged quantity language, with a long vs. short distinction for both consonants and vowels with very few syntagmatic restrictions. The opposition between long and short is possible for vowels in all syllables, stressed or unstressed, and likewise for consonants at all word internal syllable boundaries [1], [2].

Perception tests have shown that varying the $F0$ contour ("dynamic" vs. flat, high vs. low, increasing slope, shifting location of a fall, etc.) within a vowel can shift the duration boundary between long and short for many languages that have a quantity opposition (cf. e.g. [3] for a partial survey). The same has been found for Finnish [2], [4]–[6]. These perception tests have studied accented vowels (in isolated words or in carrier sentences) in word initial stressed position, a position which often exhibits an $F0$ peak in Finnish. The interaction between quantity and $F0$ in Finnish has been taken to show an affinity with tone languages, since in both cases a change in $F0$ can affect the perceived lexical identity of words [4]–[7].

Another possible interpretation is that manipulating $F0$ contour affects perceived local tempo (speaking rate). Naturally this should also affect perceived quantity, since quantity distinctions cannot be based on absolute durations, but must be judged relative to speaking rate. This type of explanation has been offered for perception data from Danish [8], Japanese [9], and Finnish [2], as well as for Finnish production data [10]. If we assume that $F0$ movements associated with sentence accent progress at a fairly constant rate for constant tempo, then more of that movement will happen during a longer vowel than during a shorter vowel. An $F0$ contour familiar to the listener will then also provide a cue for tempo—$F0$ movement progressing faster than expected will cue a faster tempo and vice versa. A local tempo model has also been suggested in connection with perception data from Estonian, Finnish, Norwegian and Italian [11]–[13], but for temporal rather than $F0$ manipulation of the context.

Any factor which affects perceived tempo in the environment of a possible quantity distinction should automatically also affect perception of that quantity. It might be expected that such an effect would be strongest for manipulations (including $F0$) during the relevant quantity domain itself, but manipulations in the immediate vicinity of the quantity being perceived might also have a sizeable effect. For instance, several studies have shown that changing context rate by means of manipulating durations can have a large impact on whether whole syllables are perceived (so-called distal speech rate effect [14]–[16]).

In the present experiment, we wanted to see whether manipulating $F0$ in the *environment* of a vowel, rather than during the vowel itself, might also shift the quantity boundary in the perception of that vowel. In addition, we wanted to investigate Finnish vowel quantity perception in unstressed, unaccented positions, where there is typically a lack of $F0$ movement.

## 2. MATERIAL AND METHODS

Two native speakers of Finnish (TN & JV, the second and third authors) were recorded producing several sets of minimal pair sentences at three different tempos. The sentence pairs differed only in the phonological length (short vs. long) of the final vowel of one word. Speakers produced the sentences two times each for three raters, first at a normal rate, then at a faster than normal rate, and finally at a slower than normal rate.

As expected, durations of carrier sentences and vowels both increase from fast to normal to slow tempo. On average it would appear that long vowels are affected by tempo changes more than short vowels, but long vowels are always longer in duration than the corresponding short vowels at all tempos, even though the ratio of long to short is reduced somewhat at faster tempos.

### 2.1. Stimuli

Two minimal pairs were selected for further use. In the first pair (speaker JV) the word (and final vowel) in question occurred at the end of the sentence:

(I) *Se on tosi ärsyttävä./ärsyttävää.*

In the second pair (speaker TN) the word and final vowel occurred in the middle of the sentence:

(II) *Se on tosi helppo/helppoo tuottaa ittekkin.*

Two tokens from each sentence pair were selected to calculate target durations for synthesizing stimuli for the perception experiment: the fast tempo sentence with the shortest long vowel and the slow tempo sentence with the longest short vowel.

A five step stimulus series was constructed for each sentence using Praat to manipulate durations in the following manner. The duration of the target vowel was kept constant at the geometric mean of the fast long vowel and the slow short vowel: 92 ms for stimulus series I (*Se on tosi ärsyttävä/ärsyttävää*), and 67 ms for stimulus series II (*Se on tosi helppo/helppoo tuottaa ittekkin*). The duration of the context (the rest of the sentence excluding the target vowel) was varied systematically so that the range between the durations of the two original contexts was equally divided into six equal steps on a logarithmic scale (i.e. the ratio of durations for adjacent steps in the series was constant). Stimuli 1–5 in the stimulus series were synthesized using the resulting central five durations. In other words, the context durations of the original sentences (not used in any stimuli) could be thought of as steps 0 and 6 in the series of equal steps. The resulting durations of target vowels and contexts are shown in Table 1 for both series, and also graphically in Figure 1 for series I.

**Table 1:** Stimulus structure

| | Duration (ms) | | | | |
|---|---|---|---|---|---|
| | series I: | | series II: | | |
| | preceding context | vowel | preceding context | vowel | following context |
| 1: | 772 | 92 | 568 | 67 | 958 |
| 2: | 874 | 92 | 624 | 67 | 1033 |
| 3: | 989 | 92 | 685 | 67 | 1114 |
| 4: | 1119 | 92 | 751 | 67 | 1202 |
| 5: | 1267 | 92 | 824 | 67 | 1296 |
| | $F0$ range (semitones) | | | | |
| A: | 14.78 | | | 11.12 | |
| B: | 7.10 | | | 5.04 | |

As expected, in spite of having the same duration, the target vowel in stimulus 1 with the shortest (fastest) carrier sentence unambiguously sounds long, whereas in stimulus 5 with the longest (slowest) carrier sentence it sounds short.

Because our main interest was to see whether $F0$ fluctuation affects quantity perception (presumably mediated by a difference in perceived tempo), we also manipulated the $F0$ contours of stimuli. For each stimulus series two versions A and B were synthesized using two different $F0$ contours. Both versions kept $F0$ of the target vowel constant (original contour). Average $F0$ during the target vowel also served as a baseline around which the $F0$ contour for the rest of the sentence was expanded or reduced. For version A, to create an expanded $F0$ range, deviation of $F0$ (in semitones) from baseline outside the target vowel was multiplied by 1.5. For version B, to create a reduced $F0$ range, deviation of $F0$ (in semitones) from baseline outside the target vowel was divided by 1.5. Thus the $F0$ ranges for the A and B versions differed from each other (in semitones) by a factor of 2.25. The resultant $F0$ ranges are shown in Table 1. Figure 2 shows the resulting $F0$ contours for series I.

One original normal tempo sentence with a long test vowel was selected for the spectral basis of the manipulated stimuli in each series. Thus no stimulus was identical to an original token, all stimuli were manipulated for both duration and $F0$. All in all a total of 20 stimuli were thus produced: 5 context duration steps (1, 2, 3, 4, 5) $\times 2$ $F0$ versions (A, B) $\times 2$ series (I, II).

### 2.2. Perception test

19 subjects took part in the perception test, 5 male, 13 female, 1 other, all university students aged 20–29 years (median age 22) at the time of the test.

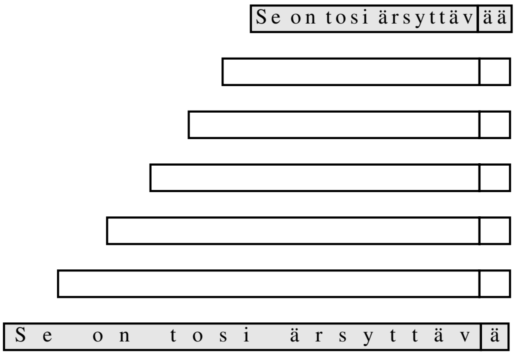


**Figure 1:** Durations for stimulus series I, steps 1–5: *Se on tosi ärsyttävä/ärsyttävää* (speaker JV, original sentences in gray)

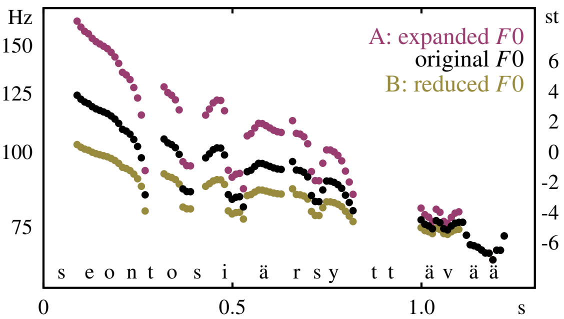


**Figure 2:** $F0$ for simulus series I, versions A & B: *Se on tosi ärsyttävä/arsyttävää* (speaker JV)

Five repetitions of each stimuli were presented to subjects in random order, with the restriction that a stimulus from the same series (sentence I or II) was never presented more than three times in a row. Subjects were required to underline on an answer sheet which of the two possible words they heard (i.e. *ärsyttävä* vs. *ärsyttävää* or *helppo* vs. *helppoo*). They were instructed to make a decision even if they were unsure which word they heard.

### 2.3. Statistical model

A Bayesian hierarchical logistic regression model [17] was used to analyze the responses of the perception test. In this model probability of a long vowel response $p$ is modeled as a logistic function of the context duration (or tempo) step $x$.

$$\text{logit}(p) = \beta \cdot (x - \mu) \quad (1)$$

Both intercept $\mu$ (50 % boundary between long and short vowel response) and slope $\beta$ are allowed to vary by main effects subject (SUBJ), stimulus series (sentence, SENT) and $F0$ version (F0), as well as all interaction effects. Slope parameters $\beta$ are restricted to be positive using a logarithmic scale.

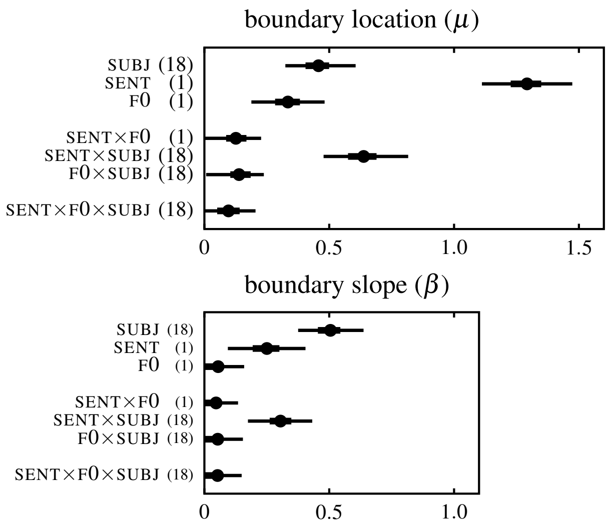


**Figure 3:** Posterior distributions (median, 50 % and 95 % HDI) for size of boundary location ($\mu$) and slope ($\beta$) effects (standard deviations of effects, degrees of freedom in parentheses)

## 3. RESULTS

Estimated (posterior) standard deviations (SD) of the effects in the model are presented visually in Figure 3. These standard deviations give an idea of the importance of each effect [17]. In this figure the posterior distributions have been stylized: the thin line shows the 95 % HDI (Highest Density Interval), the thick line shows the 50 % HDI, and the dot shows the median of the posterior distribution.

Not surprisingly, the SD for the SUBJ main effects are clearly non-zero, indicating that subjects differed among themselves, both in (50 %) boundary location $\mu$, and in slope $\beta$. Also the two sentences clearly produced different $\mu$ and $\beta$, and there was a clear SENT×SUBJ interaction. The difference in slope for the two sentences (sentence II had a slightly shallower slope) is probably due to the fact that the stimulus steps in series II were smaller because the original endpoints used to construct the stimuli were not as far apart.

The main interest for this experiment is whether expanding and reducing the $F0$ contour of the carrier had any affect on perceived quantity of the target vowel. As can be seen in Figure 3, the F0 main effect is clearly non-zero for boundary location $\mu$. The F0 main effect for slope $\beta$ is very small or zero, as are all interactions involving F0, for both $\mu$ and $\beta$. This indicates that changing context $F0$ contour (from range A to range B) shifted the identification curve towards shorter (faster) or longer (slower) contexts, but did not greatly affect its slope (or boundary width). Figure 3 does not indicate the

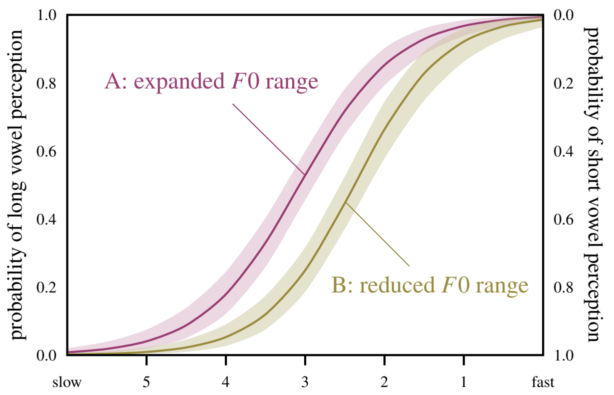


**Figure 4:** Posterior average identification curves (median and 95 % CI) for stimulus series I

direction of this shift, but looking at the posterior distributions of the individual parameters reveals that A (expanded $F0$ range) shifts the boundary towards slower context and B (reduced range) towards faster context. This difference is illustrated in Figure 4, which shows the posterior average (with all SUBJ effects removed) identification curves for stimulus series I.

The shift observed is exactly what is expected if greater $F0$ range has the effect, *ceteris paribus*, of increasing perceived tempo, because in that case it should be easier to hear a long vowel in the context of expanded $F0$ range (A).

## 4. DISCUSSION AND CONCLUSION

In a series of experiments on Estonian, Finnish, Norwegian and Italian [11]–[13] looking at the influence on quantity perception of durational variation outside the domain of quantity, Krull *et al.* concluded that "The results were compatible with a model of speech perception where an 'inner clock' handles variations in the speaking rate ..." [13]. The present results are also compatible with a local speaking rate model such as Traunmüller's modulation model [18].

Aulanko found a greater $F0$ range for Finnish long vowels compared with short vowels in the first, stressed syllable of two syllable words embedded in carrier sentences, due to the fact that although the beginning of the falling $F0$ contour was similar in both long and short vowels, it extended further downward in the long vowels. He speculated that "This might be a feature that is used unconsciously by Finns in perceiving linguistic quantity degrees: a vowel with a relatively short duration but with a great $F_0$ movement might tend to be perceived as long." [19, p. 48]. Vihanta reported similar results in a larger investigation utilizing pairs of semantically plausible sentences differing in quantity patterns. He concluded that it is probably the case that a change in $F0$, determined by sentence prosodic factors, simply progresses further during a long phoneme [10]. Suomi's production data for Finnish words of various lengths and quantity patterns embedded in carrier sentences in several accentual conditions [20], [21] showed an almost invariant $F0$ contour for each accentual condition, independent of target word length and quantity structure. All these results suggest that $F0$ contour, which is fairly constant as long as tempo and intonational structure remain unchanged, provides a robust indication of local tempo in Finnish.

It has been known for some time that changes in $F0$ (range, level, dynamics) can affect perceived quantity. Previous studies have usually manipulated $F0$ *within* the segment perceived as phonologically long or short. It now appears that changes in $F0$ *outside* that segment can affect perceived quantity as well. This means that rather than postulating various "lexical tones" associated with various quantities it is simpler, at least in the case of Finnish, to explain the effects of $F0$ on perceived quantity as mediated by tempo. In any case quantity differences must be perceived in relation to tempo, and perceived tempo (and the passage of time) is known to be affected by many factors including $F0$ movements.

The inevitable interaction between $F0$ contour and durational or rhythmic aspects of speech does, of course, mean that there can be no clear cut distinction between pure tone languages and pure quantity languages. Indeed, it is misleading to think of any language in terms of pure, mutually exclusive categories. In the course of history it is possible that a language such as Finnish might gradually come to rely more and more on local $F0$ movements to distinguish words previously distinguished mainly by durational or rhythmical means, especially if quantity distinctions came to be restricted to enviroments, such as accents, with stereotypical $F0$ movements.

However, in present day Finnish, it may be that $F0$ contours provide a very reliable cue to local tempo (thereby also influencing quantity) precisely because Finnish does not have lexical tone. Rather, $F0$ "is tied more closely to a different level of timing (eg. sentence intonation), and may thus provide a reference for perception of quantity." [2]. Put in a different way, Finnish quantity is not tonal, rather tonality (intonation) provides the listener with a strong cue for synchronizing to the time scale of the speaker at the level of quantity.

## 5. REFERENCES


[1] J. Lehtonen, *Aspects of Quantity in Standard Finnish* (Studia Philologica Jyväskyläensia VI). University of Jyväskylä, 1970.

[2] M. O'Dell, *Intrinsic Timing and Quantity in Finnish* (Acta Universitatis Tamperensis 979). Tampere University Press, 2003.

[3] R. Cumming, "The effect of dynamic fundamental frequency on the perception of duration," *Journal of Phonetics*, vol. 39, pp. 375–387, 2011.

[4] M. Vainio, D. Aalto, J. Järvikivi, and A. Suni, "Quantity and tone in Finnish lexically stressed syllables," in *Proceedings of the 2nd International Symposium on Tonal Aspects of Languages*, A. Belotel-Grenié and M. Grenié, Eds., 2006, pp. 121–124.

[5] J. Järvikivi, D. Aalto, R. Aulanko, and M. Vainio, "Perception of vowel length: Tonality cues categorization even in a quantity language," in *Proceedings of the 16th International Congress of Phonetic Sciences*, J. Trouvain and W. J. Barry, Eds., Universität des Saarlandes, Saarbrücken, Germany, 2007, pp. 693–696.

[6] J. Järvikivi, M. Vainio, and D. Aalto, "Real-time correlates of phonological quantity reveal unity of tonal and non-tonal languages," *PLoS ONE*, vol. 5, no. 9, e12603, 2010.

[7] J. Šimko, D. Aalto, P. Lippus, M. Włodarczak, and M. Vainio, "Pitch, perceived duration and auditory biases: Comparison among languages," in *Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK*, The Scottish Consortium for ICPhS 2015, Ed., University of Glasgow, 2015, paper 575, ISBN: 978-0-85261-941-4.

[8] N. Thorsen, "$F_0$ timing in Danish word peception," *Phonetica*, vol. 41, pp. 17–30, 1984.

[9] Y. Nagano-Madsen, *Mora and Prosodic Coordination: A Phonetic Study of Japanese, Eskimo and Yoruba*. Lund University Press, 1992.

[10] V. V. Vihanta, "$\mathrm{F}_0$:n osuudesta suomen kvantiteettioppositiossa," in *Fonetiikan päivät — Espoo 1988 / Papers from the 15th Meeting of Finnish Phoneticians*, M. Karjalainen and U. K. Laine, Eds., ser. Teknillinen Korkeakoulu, Sähkötekniikan osasto, Akustiikka ja äänenkäsittely / Helsinki University of Technology, Faculty of Electrical Engineering, Acoustics Laboratory, (résumé: Sur le role de la $\mathrm{F}_0$ dans l'opposition de quantité en finnois), Helsinki University of Technology, 1988, pp. 13–37.

[11] D. Krull, H. Traunmüller, and W. A. van Dommelen, "The effect of local speaking rate on perceived quantity: A comparison between three languages," in *Proceedings of the 15th International Congress of Phonetic Sciences*, M. J. Solé, D. Recasens, and J. Romero, Eds., Universitat Autònoma de Barcelona, Spain, 2003, pp. 1739–1742.

[12] H. Traunmüller and D. Krull, "The effect of local speaking rate on the perception of quantity in Estonian," *Phonetica*, vol. 60, pp. 187–207, 2003.

[13] D. Krull, H. Traunmüller, and P. M. Bertinetto, "Local speaking rate and perceived quantity: An experiment with Italian listeners," *Working Papers (Lund University, Dept. of Linguistics)*, vol. 52, pp. 81–84, 2006. [Online]. Available: https://www2.ling.su.se/staff/hartmut/quantity_italian.pdf.

[14] L. C. Dilley and M. A. Pitt, "Altering context speech rate can cause words to appear or disappear," *Psychological Science*, vol. 21, no. 11, pp. 1664–1670, 2010. [Online]. Available: http://journals.sagepub.com/doi/pdf/10.1177/0956797610384743.

[15] L. C. Dilley, T. H. Morrill, and E. Banzina, "New tests of the distal speech rate effect: Examining cross-linguistic generalization," *Frontiers in Psychology*, vol. 4, no. 1002, 2013. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fpsyg.2013.01002/full#B18a.

[16] M. O'Dell and T. Nieminen, "Distal rate effect for Finnish epenthetic vowels," in *Proceedings 9th International Conference on Speech Prosody 2018*, 2018, pp. 646–650. DOI: 10.21437/SpeechProsody.2018-131.

[17] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2007.

[18] H. Traunmüller, "Conventional, biological and environmental factors in speech communication: A modulation theory," *Phonetica*, vol. 51, no. 1–3, pp. 170–183, 1994.

[19] R. Aulanko, "Microprosodic features in speech: Experiments on Finnish," in *XIII Fonetiikan päivät Turku 1985 / XIII Meeting of Finnish Phoneticians — Turku 1985*, O. Aaltonen and T. Hulkko, Eds., ser. Publications of the Department of Finnish and General Linguistics of the University of Turku, 1985, pp. 33–54.

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

[21] K. Suomi, "On the tonal and temporal domains of accent in Finnish," *Journal of Phonetics*, vol. 35, no. 1, pp. 40–55, 2007.