ON PITCH AND PERCEPTUAL PROMINENCE IN CONVERSATIONAL FINNISH SPEECH

Mietta Lennes

University of Helsinki
mietta.lennes@helsinki.fi

ABSTRACT

For the Finnish language, the tonal correlates of sentence accent or word-internal stress patterns have only been studied in read-aloud laboratory speech. In this preliminary study, the general relationship between pitch patterns and perceived prominence of word-initial syllables is investigated in free conversational Finnish for one female and one male speaker. The typical pitch levels and distributions are also compared across speakers.

Keywords: pitch, F0, prominence, conversational speech, Finnish, dialogue

1. INTRODUCTION

In speech, some words, syllables or other portions of speech are perceived as more *prominent* than others, i.e., the prominent portions "stick out" from their environment. Prominence is considered as a perceptual and gradual property. Functionally, however, prominence can be seen to result from, e.g., sentence accent and/or word stress.

In addition to syntactic and semantic factors, several acoustic-phonetic parameters, e.g., pitch, intensity, and segmental durations, may be associated with perceived prominence to varying degrees. Pitch patterns have been found to be an important prominence cue in several languages, such as Dutch [3] and Finnish. However, the tonal correlates for prominence in Finnish have only been studied for read-aloud speech [6, 5, 7].

The Finnish language does not exhibit lexically distinctive word stress, but when a word is clearly pronounced or accented in speech, the word-initial syllable is usually perceived as the most prominent one within the word [4]. For isolated, read-aloud Finnish sentences, the accentuation of a word has been found to be associated with a rise-fall pitch pattern where a higher pitch level and a steeper pitch fall tend to increase the degree of perceived prominence, although the interpretations of the domain of this prominence cue vary [6, 5, 7]. The present preliminary study asks what kind of pitch patterns are typical for prominent words in free conversational Finnish speech.

2. Material

Informal unscripted dialogues were recorded from ten young Finnish adults (five females). The participants in each dialogue were close friends and they were allowed to chat freely and unmonitored in an anechoic room for a total of 40 to 60 minutes on either given or self-selected topics. The speakers were sitting a few meters apart and facing opposite directions. Each speaker's speech was recorded to a separate channel of a DAT recorder using high-quality headset microphones. The recorded material was then transferred to a computer and sampled to the precision of 22,05 kHz. The two channels of the stereo files were separated, resulting in one audio file per speaker.

Each speaker's utterances were delineated and orthographically transcribed using the Praat program [1]. Parts of the material were phonetically segmented and transcribed, and word and syllable boundaries were marked.

2.1. Prominence marking

Prominent syllables were marked in the speech of one female (age 24 years) and one male Finnish speaker (age 28 years) by one trained phonetician. Prior to the prominence judgments, all word-initial syllables were automatically marked on the basis of the annotations. Each utterance (a stretch of speech demarcated by pauses) was then played back several times and the syllables within that utterance were labeled according to their perceived prominence. This judgment was performed solely on an auditory basis, without inspecting acoustic displays. Those syllables that were heard as clearly prominent with respect to the rest of the utterance were labeled with the number 2, possibly but not clearly prominent syllables received the label 1, and those syllables that were not at all prominent were left unlabeled. This procedure was repeated for all utterances, irrespective of their length or type.

The number of syllables analyzed for each speaker is shown in table 1. Approximately 23 % of the word-initial syllables of the female speaker and 25 % of the syllables of the male speaker were judged as clearly prominent, i.e., according to this listener, approximately one in every four words was prominent to some extent.

Table 1: Number of word-initial syllables analyzed for the female speaker F1 and the male speaker M1.

	F1	M1
Nonprominent	672	335
Uncertain	35	5
Prominent	215	115
Total	922	455

2.2. Pitch analysis

In order to inspect overall pitch distributions, pitch values were collected at 20 ms time steps within all the annotated utterances for ten speakers (five females). The pitch analysis was performed using the standard algorithm available in the Praat program [1]. The maximum and minimum frequency parameters required for pitch detection were separately defined for each speaker. A total of 42158–73358 pitch points were obtained for each speaker. For each individual pitch value, the relative time point within the utterance was recorded, the time 0 denoting the starting point of the utterance, and the time 1 referring to the end time of the utterance. For comparing prominent and non-prominent words, the pitch value at the temporal midpoint of each word-initial syllable and the corresponding values for its preceding and following syllable were also recorded.

In Finnish, many speakers tend to use a creaky voice quality, especially near the end of utterances. Creak is often reflected as irregular periodicity in the speech signal. In such cases, the pitch analysis algorithms tend to produce downward “octave jumps”, which cannot be consistently interpreted. When inspecting the overall distributions of sampled pitch values for speakers of Finnish, a majority of the values tend to accumulate around the speaker’s mean or median pitch. Due to creaky voice, however, a distinct but much smaller heap of pitch values can usually be observed below the major pitch cluster. For the present study, the upper boundary for this small cluster was first determined by visually inspecting the pitch distribution for each individual speaker, and these low pitch values due to creaky voicing were then excluded from further analyses.

2.3. Rescaling the pitch values

In order to be able to assess the perceptual importance of pitch changes, the semitone scale (re 100 Hz) was used for measuring pitch for the present analysis. The pitch distributions for ten speakers (five females) are shown in the upper part of figure 1. One male speaker had a higher voice than the other males, and thus a majority of his pitch values overlap with the typical distributions for female speakers. However, the shape of the distributions looks remarkably similar for all speakers.

It may be hypothesized that listeners are able to adjust to the ”typical pitch level” of a certain speaker, and that large deviations from this typical pitch might gain the listener’s attention and thus be more likely to be perceived as prominent. Since it was the intention of the present study to be able to compare relative pitch levels for two speakers, one female and one male, it was necessary to rescale the pitch values without affecting the shape of the individual pitch distributions. This was done by referring the pitch values for each speaker to their individual pitch mode, determined as the location of the maximum in the estimated pitch density function:

(1) $Mo_s = \mathrm{argmax}\, d(f_0)$

where Mo_s represents the mode pitch for speaker s and $d(f_0)$ represents the density function d over pitch f_0. The resulting distributions are shown in the lower part of figure 1. This rescaling method brings the pitch distributions together and allows for comparisons across speakers, although the shape of each distribution remains intact. Thus, the relative pitch value zero corresponds to the “preferred” pitch value for all speakers.

The results suggest that different speakers may in fact have perceptually similar pitch ranges that they prefer to use in their speech. However, speakers may also exhibit individual deviations in how they use this typical range. For instance, the figure 1 displays slight irregularities in the pitch distributions for two female speakers who apparently used their higher pitch levels slightly more than others. Moreover, by changing the mode of laryngeal vibration, it is possible to produce creak, falsetto, and other effects that are not captured by the present analysis.

2.4. Pitch distribution within utterances

At the beginnings of utterances, speakers tend to produce higher pitch levels than utterance-finally, which is probably due to physiological reasons. The downward trend in pitch movement is often referred to as declination [2]. Apparently, in order for syllables to be perceived as prominent at the beginning of utterances, they need to be produced at a higher pitch level than syllables that occur later in the utterance. Therefore, it was necessary to control for the utterance-internal position of pitch values that were to be compared. A second-order linear model was

Figure 1: Pitch densities in conversational Finnish speech for ten native young adults (age 20-30 years). Solid lines indicate distributions for female speakers, dashed lines for male speakers. The upper figure shows the pitch distributions in absolute values, the lower figure with reference to each speaker's pitch mode.

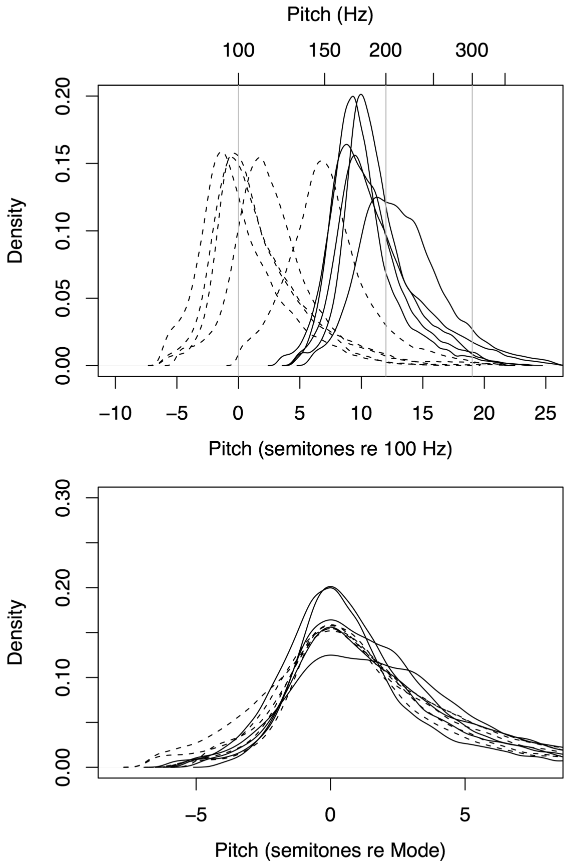
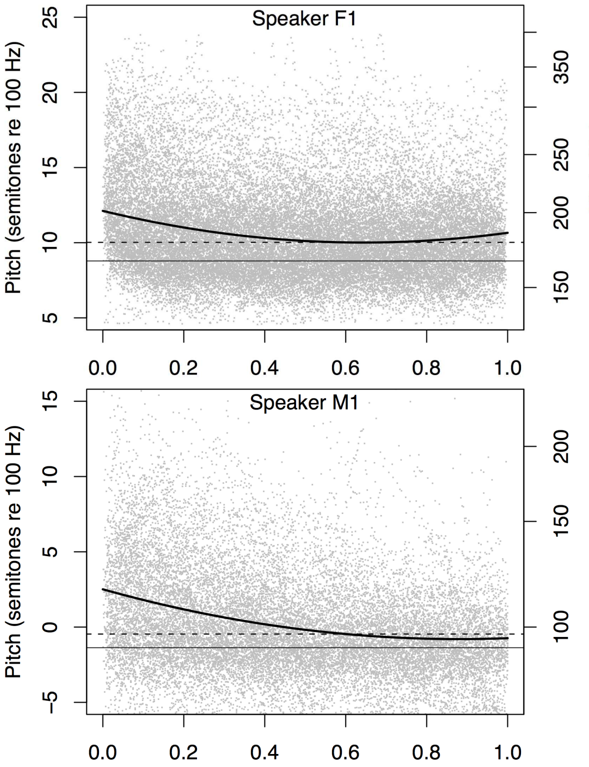


Figure 2: Pitch patterns within time-normalized utterances of one female speaker (F1) and one male speaker (M1). The horizontal axis depicts relative time within each utterance. The thick solid black curve indicates the predicted values for a second-order linear model of pitch over time. The overall median pitch is indicated with a horizontal dashed line. The overall pitch mode (the location of the maximum in the estimated density function of all pitch values) is shown with a horizontal solid line.



calculated for summarizing the pitch trend over relative time within utterances. These trendlines are shown as black curves in Fig. 2.

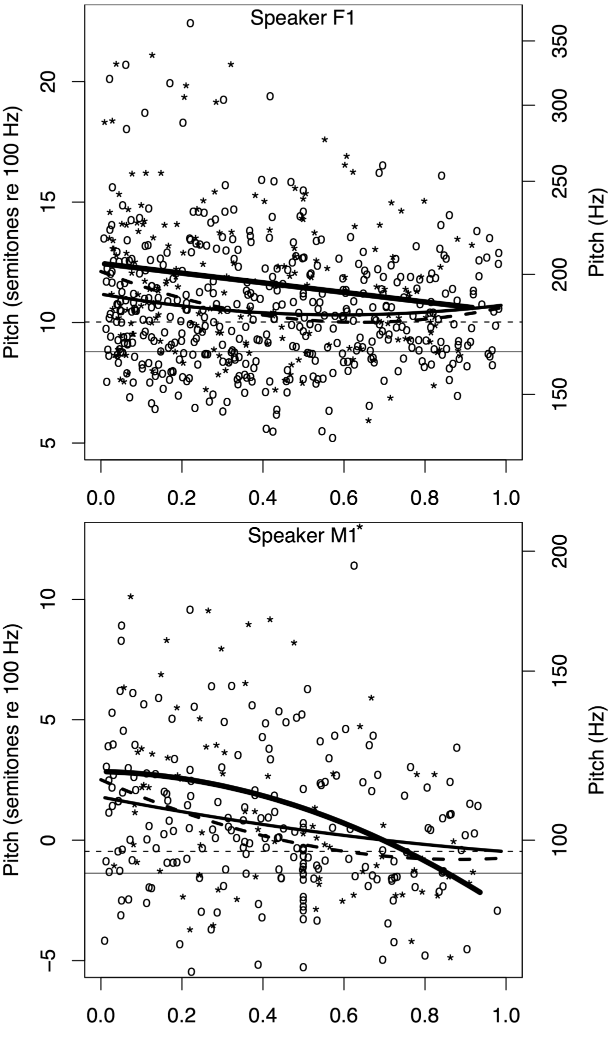
3. RESULTS AND DISCUSSION

The pitch levels and peaks or troughs within clearly prominent and clearly non-prominent word-initial syllables were compared. According to a two-sample t-test, the pitch values (in semitones referred to speaker mode) taken from the temporal mid points of word-initial syllables were significantly different ($p < 0.0001$) for prominent (median 2.5 ST) and non-prominent (median 1.6 ST) words. A similar significance level was obtained for the pitch maxima within the syllables (3.3 and 2.4 ST). However, as might be expected, the minimum pitch values within the syllables were similar for prominent and non-prominent syllables (median 0.5 ST for both). Thus, the peak height at word-initial syllables has at least some importance for prominence perception in conversational Finnish.

Figure 3 shows the individual absolute pitch values in semitones for the mid points of prominent vs. non-prominent syllables for the two speakers. In addition, second-order linear models are shown for the temporal midpoint pitch values of prominent (thick solid curve) and non-prominent (thin solid curve) syllables as well as for the general pitch contours (dashed curve). The model curve for non-prominent syllables is rather close to the overall trendline, whereas the prominent syllables exhibit slightly elevated pitch levels, especially around the middle of utterances.

The pitch difference between maximum and minimum values within each syllable only showed a significant correlation with perceived prominence for one of the speakers. However, this was probably due to the somewhat unsymmetric pitch distributions.

Figure 3: Distribution of pitch values at the temporal midpoints of word-initial syllables for prominent (*) and non-prominent (o) syllables for one female speaker (upper figure) and one male speaker (lower figure). The horizontal axis depicts relative time within each utterance. The thick solid black curve indicates a second-order linear model describing pitch levels for prominent syllables, and the thin solid black curve shows a corresponding model for non-prominent syllables. The thick dashed curve represents a model for all pitch values. The overall median pitch is indicated with a horizontal dashed grey line and the overall pitch mode with a solid grey line.



Since the median differences were 2.4 ST within prominent and 1.5 ST within non-prominent syllables, a difference in the magnitude of pitch changes probably exists.

Lastly, the pitch values in each word-initial syllable were compared with those obtained from the preceding and following syllable. The median of the pitch difference from the midpoint of the preceding syllable to the midpoint of each word-initial syllable was 0.7 ST for prominent and -0.3 ST for non-prominent syllables, and this difference was significant ($p < 0.001$). However, the difference for the pitch change from the word-initial syllable to the following syllable was even more significant ($p < 0.00001$) with a median decrease of 1.5 ST after prominent syllables and no increase or decrease at all for non-prominent syllables.

4. CONCLUSIONS

In this preliminary study, the relationship between perceived prominence and the pitch patterns around word-initial syllables was investigated for the conversational Finnish speech of two speakers. The pitch level was found to be slightly higher for prominent syllables, and a pitch rise with respect to the preceding syllable also occurred more often for prominent than for non-prominent word-initial syllables. The strongest correlation with prominence was found for the magnitude of pitch decrease from the prominent (word-initial) syllable towards the next syllable. These results are in accord with earlier studies for read-aloud Finnish. However, in some cases, prominent syllables did not follow the aforementioned pitch pattern, suggesting that other cues for prominence may have taken over.

It is likely that only the most general tonal correlates of prominence were captured by this analysis. Syllable type and syntactic or semantic factors were not considered, although both are surely important variables for perceptual prominence and may interact with pitch patterns. Also, the speech of only two speakers was judged by only one labeler. More work is thus required in order to generalize and elaborate the results.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2006. Praat: doing phonetics by computer (Version 4.3.23) [Computer program].

[2] Cohen, A., ’t Hart, J. 1967. On the anatomy of intonation. *Lingua* 19, 177–92.

[3] Streefkerk, B. M. 2002. *Prominence. Acoustic and lexical/syntactic correlates.* PhD thesis University of Amsterdam.

[4] Suomi, K. 2005. Suomen kielen prominenssien foneettisesta toteutumisesta. *Virittäjä* 109, 221–243.

[5] Suomi, K. 2007. On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics* 35, 40–55.

[6] Suomi, K., Toivanen, J., Ylitalo, R. 2003. Durational and tonal correlates of accent in Finnish. *Journal of Phonetics* 31, 113–138.

[7] Vainio, M., Järvikivi, J. 2006. Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics* 34, 319–342.