# Focus in production: Tonal shape, intensity and word order

**Martti Vainio and Juhani Järvikivi**
*Department of Speech Sciences, University of Helsinki, P.O. Box 9 (Siltavuorenpenger 20), FIN-00014, Helsinki, Finland and Department of Psychology, University of Turku, Assistentinkatu 7, FIN-20014 Turku, Finland*
*martti.vainio@helsinki.fi, juhani.jarvikivi@utu.fi*

**Abstract:** The effect of word order and prosodic focus on the tonal shape and intensity in the production of prosody was studied. The results show that the production of focus in Finnish follows a global pattern with regard to tonal features. The relative pitch height difference between contrasted words is the most important pitch-related factor in signaling narrow prosodic focus. Narrow focus is not localized to prosodically emphasized words only but relates to the utterance as a whole. It was also found that syntactic structure with respect to both intensity and tonal structure modulated relative prosodic prominence of individual words.


**PACS numbers:** 43.70.Fq, 43.70.−h, 43.70.Bk [AL]
**Date Received:** August 14, 2006 **Date Accepted:** November 26, 2006

## 1. Introduction

The question investigated in the present study is whether and how prosody and syntactic structure interact in the production of information structure, particularly in the production of focus. Focusing is used to draw attention, to contrast or to emphasize the importance of a particular part of an utterance. In general, focus is signaled by the speaker by making one or another part of an utterance prominent either syntactically or prosodically. Focus, as understood here, pertains to new propositional information evoked by the focused element and what is pragmatically presupposed by the speaker and the hearer, i.e., to information which is not already presupposed or talked about, and thus not shared by the speaker and hearer. In this sense, focus is seen as an abstract proposition emerging from the combination of the presupposed or shared information and a sentence element that is marked as more prominent by prosody or syntax. In the present study the relation between syntactic and prosodic prominence and information structure in Finnish is exploited in investigating whether changes in syntactic structure together with different propositions are reflected in prosodic parameters of voice fundamental frequency ($f_0$) and intensity in production.

Finnish is an agglutinative-fusional language with flexible word order. The word order in Finnish frequently serves to signal information structure. For example, in an unmarked case, such as (1) "Menemme laivalla Jimille" (we go by boat to Jimi's), the canonical order of the two adverbs laiva+lla (boat+with) and Jimi+lle (Jimi+to) (i.e., manner+place) conforms to its default information structure. Consequently, no propositions over and above what is already explicitly asserted is evoked by the word order. In contrast, changing the word order to marked (2) "Menemme Jimille laivalla" highlights the last element laivalla (with boat). Therefore, rather than just asserting that we are in fact going to Jimi's, this presupposed information together with the marked position of "laivalla" evokes the proposition that it is by boat we are going to Jimi's and not by a car, for example—as if it were an answer to a question "how do you go to Jimi's?." For the pragmatic use of word order in Finnish, see, e.g., Ref. 1.

In addition to word order, prosody can be used to mark any constituent under the domain of focus even in the syntactically unmarked case by increasing the accent or stress on the part of an utterance that is intended to be brought into focus. Thus, a Finnish speaker can say "Menemme *Jimille* laivalla" as well as "Menemme Jimille *laivalla*" (italics depict prosodic focus). An important question is, then, whether and how the two main means available for

signaling focus in Finnish—syntactic and prosodic—interact in production when one or another part of an utterance needs to be prosodically marked as focused.

Prosodic (narrow) focus is usually achieved by the speaker by increasing the prominence of the focused constituent in the utterance. This is usually done by making the local $f_0$ excursions bigger and attenuating others. Usually these are accompanied by respective increases and decreases in intensity as well as segmental durations. The $f_0$ and intensity changes do not, however, correlate perfectly, and their interactions tend to be fairly complex. Thus, the perception of prominence is tightly coupled with the perception of pitch in speech. For instance, Pierrehumbert[2] showed that a later $f_0$ peak in an utterance has to be lower than the previous ones to be perceived as having an equally high pitch in English. This has since been shown to hold in many other languages (for instance Dutch[3] and Finnish[4]). The situation in tone languages seems to be more complex, although declination has been attested for at least Mandarin Chinese.[5]

Vainio and Järvikivi[4] studied the role of intensity and accentuation in the perception of prominence in Finnish. On one hand, they found that *ceteris paribus* a word order reversal had an effect on the perceived relative prominence of two words in a short Finnish utterance (verb followed by two nouns inflected to act as adverbials of place and manner; we will use only the term "noun" from now on except when we refer to both of them at once as an adverbial phrase). They explained this finding to reflect the fact that, since the word order reversal resulted in the latter word being syntactically marked for focus, the participants perceived it as being also prosodically more prominent, despite the fact that pitch and intensity were controlled with respect to the unmarked word order condition. On the other hand, they also found that the prominence of the two nouns in the utterance followed a so-called *flat-hat* pattern; i.e., the prominence of the earlier word related to the $f_0$ rise and the prominence of the latter words was related to the $f_0$ fall, with the relative heights of the peaks being the most important factor when subjects were asked to indicate which word (if any) they perceived as the most prominent. In other words, the fall of the earlier peak and the rise of the later peak (i.e., the transition between the peaks) did not contribute significantly to the perception of prominence of either peak, which both exhibit characteristic pointed-hat patterns.

We designed a production experiment to test hypotheses formed according to the findings above: (1) If the perceptual bias in the Vainio and Järvikivi study was caused by the change in information structure evoked by the change in word order, the same phenomenon should be reflected in production as compensation with regard to prominence, which should be significantly less pronounced with marked word order that already syntactically focuses one of the critical words, and (2) the different narrow focus conditions should form a pattern where the difference between the $f_0$ peaks (usually referred to as top-line declination) is the most important contributing factor together with a rise (in the case of "early focus") and a fall (in the case of "late focus"). It is within these variables where the word order-induced compensation should be visible.

## 2. Materials

A simple declarative sentence starting with a verb and ending with an adverbial phrase whose word order could be reversed to mark the sentence for focus was used; the basic sentence "Menemme laivalla Jimille" (go-we boat-by Jimi's-to) or with a reversed word order "Menemme Jimille laivalla" (go-we Jimi's-to boat-by), allows for three different focus conditions with regard to the nouns *laiva*, and *Jimi*: broad focus (no specific prosodic marking for emphasis), narrow focus on the first noun, and narrow focus on the second noun. Two different words were used for the vehicle ["laiva" (boat) and "juna" (train)] and three proper nouns for the person to be visited (Jimi, Jani, and Lumi). With three different focus conditions and two different word order conditions, a set of 36 different sentences was created. Accordingly, a set of prompt questions matching the intended three focus conditions was created as follows: Broad focus; Mitä teette tänään (what do you do today)?, Narrow focus on "laivalla;" Millä menette Lumille (with what/How do you go to Lumi's)?, and Narrow focus on "lumille;" Minne menette laivalla (where do you go by boat)?. The question prompts were then recorded by a female speaker to be

 

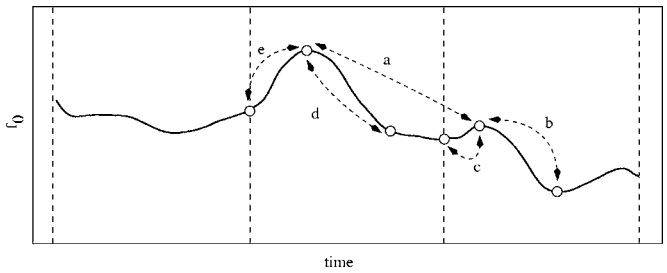


Fig. 1. A schematic illustration of the measured differences on an $f_0$ contour. Vertical lines depict word boundaries, and dotted lines with arrows marked with letters (a–e) show the differences (calculated in semitones) used in analyses. The pitch contour is a time-normalized average of the broad focus items used in the analyses.

presented to the participants in order to elicit the desired prosodic focus in the reply which was read from a sheet of paper.

*2.1 Participants and procedure*

Eight participants (seven female) took part in the experiment. All of the participants were choir members living in the Helsinki area with similar backgrounds in eastern Finland. None of the participants was familiar with speech research and none reported any hearing problems. All of the speakers spoke with a neutral Helsinki area dialect/accent.

The 36 prompt-reply pairs were randomized for each participant and he or she was given a sheet of paper with the corresponding replies. The focus was not indicated in any way on the paper as it was intended to be elicited by the type of question to be presented to the participants. The participants were not informed of the nature of the experiment and were asked to speak lively.

The prompt questions were played to the participants through a high-quality loudspeaker (Genelec 1029A) in a sound-treated recording studio at the Department of Speech Sciences at the University of Helsinki. The prompts were spaced so that the participants had ample time to reply. The replies uttered by the participants were recorded directly to a computer hard disk at 44.1-kHz sampling frequency and 16-bit quantization using a high-quality analog-to-digital converter (Digi002 by Digidesign) and a high-quality condenser microphone (AKG 4000B).

## 3. Results

Before data analyses each of the participants' responses was labeled and both intensity and $f_0$ were calculated using the PRAAT program.[6] The utterances were then annotated manually. Three points of interest for each word in the utterance were marked on the $f_0$ curve and the segmental contents were labeled on a syllable basis. The three points for a given word corresponded to the basic pointed-hat pattern mainly used for accentuation in Finnish: the first point corresponding to the start of the $f_0$ rise, the second point to the peak, and the last point to the end of the $f_0$ fall. Both $f_0$ and intensity were measured at these points for subsequent statistical analyses. An example of the analysis points can be seen in Fig. 1.

Logistic regression analyses were conducted to determine which pitch-related factors contributed to the pitch contours' belonging to a given focus category. Several regression models were estimated using the pitch differences relevant to the formulated predictions. All models were cross validated (40 repetitions) and backwards elimination was used to determine the significant predictors. The predictors were also tested for nonlinearities using restricted cubic splines[7] (all of the predictors turned out to behave in a linear fashion). We also tested the interactions between the predictors.

Before the analyses were conducted, the first author marked all utterances considered problematic with regard to $f_0$ patterns by visually inspecting the curves. In total, 36 utterances were identified as problematic and were played—together with the same number of filler

  

Table 1. Pitch differences semitones as used in the regression analyses: Means (upper part) and standard deviations (lower part).

| Condition | peakdif a | lfall b | lrise c | ffall d | frise e |
|---|---|---|---|---|---|
| B | 3.193 | 3.806 | −1.179 | 3.311 | −3.258 |
| N1 | 6.928 | 1.406 | −0.805 | 6.220 | −3.784 |
| N2 | −1.437 | 6.275 | −2.345 | 1.352 | −2.275 |
| B | 2.139 | 2.399 | 1.493 | 2.392 | 2.463 |
| N1 | 2.428 | 1.789 | 1.095 | 2.195 | 2.463 |
| N2 | 2.215 | 2.067 | 1.169 | 1.337 | 1.581 |

utterances—to a group of 20 naive listeners who judged the focus condition of each utterance. Utterances whose focus was judged to be the intended one by fewer than four listeners were removed from the regression analyses as outliers. All in all, 12 trials used in the analyses were rejected this way (4% of the data). (Note that for the ANOVAs a different method of removing outliers was used.) The rises, falls, and peak height differences used in the statistical analyses were calculated in semitones. The intensities were calculated in decibels; the values were measured instantaneously at the peak $f_0$ points.

*3.1 Tonal pattern*

The following predictions concerning the tonal pattern of accentuation were tested: (1) the most important feature responsible for the focus conditions is the difference in peak heights of the two accented words (a in Fig. 1); (2) the rise of the first peak (line e) is more important than the fall (d in Fig. 1); and (3) the fall of the latter peak (b in Fig. 1) is more important than its rise (c in Fig. 1). The relative importance of the features was tested with logistic regression, using the different pitch-related features as predictors and the given focus condition as the dependent variable. Analysis was only performed on the two narrow focus conditions as there were no predictions with regard to the broad focus condition. The measurements (marked with letters from a to e) can be seen in Fig. 1. The semitone values are summarized in Table 1. The analyses of the narrow focus conditions are discussed separately below.

3.1.1 Early focus

The early focus (condition N1) was best explained by a model that included only the peak height difference (peakdiff a), the last peak fall (lfall b), and their interaction. The most important predictor was the difference between the peak heights $[\chi^2(2)=42.99, p<0.0001]$, followed by the last fall $[\chi^2(1)=24.15, p<0.0001]$ and their interaction $[\chi^2(1)=20.28, p<0.0001]$. The whole model was naturally highly significant $[\chi^2(3)=74.55, p<0.0001, R^2(3)=0.735]$. That is, more than 70% of the categorization can be accounted for by pitch alone.

3.1.2 Late focus

The late focus condition (N2) turned out to be more complex requiring two additional predictors, mainly the ones describing the first peak (ffall d, and frise e). The main predictor was again the peak height difference $[\chi^2(2)=24.07, p<0.0001]$, followed by the last fall $[\chi^2(2)=10.24, p<0.0060]$, and their interaction $[\chi^2(1)=8.41, p<0.0037]$. The first peak rise and fall were also highly significant $[\chi^2(1)=8.44, p<0.0037]$ and $[\chi^2(1)=6.41, p<0.0113]$, respectively. The overall model was again highly significant $[\chi^2(5)=44.46, p<0.0001, R^2(5)=0.833]$.

Figure 2 shows the $f_0$ contours in all focus conditions averaged over all speakers. The $f_0$ contours can be summarized in terms of pitch range adjustments from the broad focus in order

 

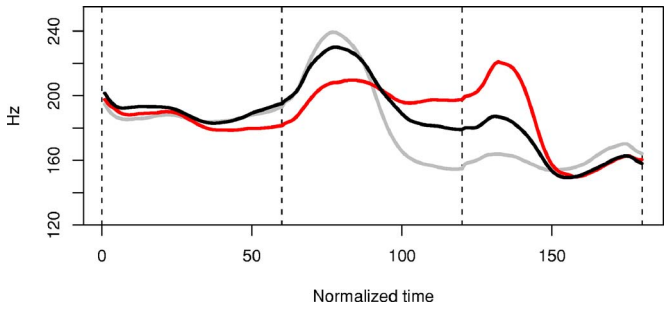


Fig. 2. (Color online) $f_0$ contours averaged over all focus conditions: broad focus; black line, narrow focus on first noun; light gray line, narrow focus on last word; red line. Word boundaries are depicted by vertical dashed lines. Each word was time normalized and averaged separately using 60 equidistant time points.

to maximize the difference between the peak heights without deaccenting the nonfocused word. That is, the basic tonal shape is preserved. Similar results have been found for English by Xu and Xu.[8]

### *3.2 Word order*

To investigate whether the manipulation of word order interacted with intensity and tonal shape (pitch) in the production of focus, $2 \times 3$ analyses of variance (ANOVAs) with word order (marked, unmarked) and focus (broad, N1, N2) as factors were done with peak difference in decibels (intensity) and semitones (pitch) as the dependent measure. Both by-subject (participant means averaged over items—$F1$) and by-item analyses (item means averaged over participants—$F2$) are reported. Before analyzing the data for pitch, outliers were removed using 2.5 SDs below and above the condition means as a criterion. The outliers in the pitch experiment accounted for 2.8% of all data points. As these outliers were mostly caused by production errors, i.e., the speaker either interpreting the question incorrectly or producing a clearly unintended response, these outliers were also removed from the analyses of intensity. Additionally, all further data points 2.5 SDs over or above the condition means were removed from the analyses of intensity. These outliers accounted for a further 4.2% of the data. The condition means for intensity and pitch are summarized in Table 2. The results from all subsequent statistical analyses are given in Table 3.

In both analyses (pitch and intensity) word order (unmarked, marked) and focus (broad=B, noun 1=Nl, noun 2=N2) were within variables in the subject analyses ($F1$). In the item analyses ($F2$) word order was a between-item factor. In what follows we discuss the results of each of the analyses separately.

#### 3.2.1 Intensity

As expected there was a significant main effect of focus. In addition, however, the results also showed a significant interaction between word order and intended focus. Further pairwise com-

Table 2. Peak differences in decibels (intensity) and semitones (pitch) for the emphasized and nonemphasized prompts in the marked and unmarked word order conditions for broad focus (B), narrow focus on the first word (N1), and narrow focus on the second word (N2).

| | Pitch | | Intensity | |
|---|---|---|---|---|
| Condition | Unmarked | Marked | Unmarked | Marked |
| B | 2.68 (2.00) | 4.06 (2.52) | 2.25 (2.21) | 3.83 (2.26) |
| N1 | 7.01 (1.99) | 7.17 (2.14) | 7.50 (3.92) | 6.95 (4.32) |
| N2 | −1.25 (1.94) | −1.58 (2.00) | 0.40 (3.67) | 0.50 (3.71) |

  

Table 3. Results from the overall analyses of variance with decibels (intensity) and semitones (tonality) as the dependent measures. Statistical significances are marked as follows: $(*)p<0.1$,$^{*}p<0.05$,$^{**}p<0.01$,$^{***}p<0.001$.

| | Pitch | | Intensity | |
|---|---|---|---|---|
| Condition | *F*1 | *F*2 | *F*1 | *F*2 |
| WO. | 3.768 $(^{*})$ | 1.116 | <1 | <1 |
| Foc | $103,62^{***}$ | $423.29^{***}$ | $31.29^{**}$ | $203.70^{***}$ |
| WO×Foc | $4.981^{*}$ | $4.595^{*}$ | $5.830^{*}$ | 3.645 $(^{*})$ |

parisons showed that the broad focus condition was significantly modulated by whether the word order was marked or not $[t1(7)=3.375, P=0.012; t2(5)=2.706, p=0.042]$.

### 3.2.2 Pitch

Again the results showed a significant main effect of focus. Additionally, this effect was further qualified by a significant interaction between focus and word order. Pairwise contrasts showed that the interaction was mainly due to a significant difference between the marked and unmarked broad focus conditions $[t1(7)=3.062, p=0.018; t2(5)=2.274, p=0.072]$, although the difference was not statistically significant, most likely due to a lack of power; the tendency to compensate for the word order reversal can be seen numerically in the N2 condition as well $(t's<1.12, p's>0.15)$ as well as graphically in Fig. 3.

## 4. Discussion

The results presented in this paper reveal a complicated phenomenon relating to the production of focus in Finnish.[4] First, the results regarding the overall tonal shape are in consonance with results on perception of prominence. That is, the production of narrow focus in Finnish follows a flat-hat pattern with regard to the tonal features used for increasing local prominence; a fall in the final word of an utterance and a rise on a nonfinal word. However, the patterns are somewhat more complex due to the fact that the participants mostly produced patterns with a so-called sagging transition (a clear valley between the two peaks) between the two peaks, as a real flat-hat pattern would signify a different pragmatic meaning in Finnish. More interestingly, the pattern for the early focus only depends on the peak height difference and the fall of the latter, final accent. It seems that in this case the speakers are controlling the pitch features in a holistic manner and concentrate on the attenuation of parts more than on the intensification. In summary, the production of prosodic focus is not localized to the prominent or emphasized word only but relates to the time domain of the whole utterance or at least to the part of it where the

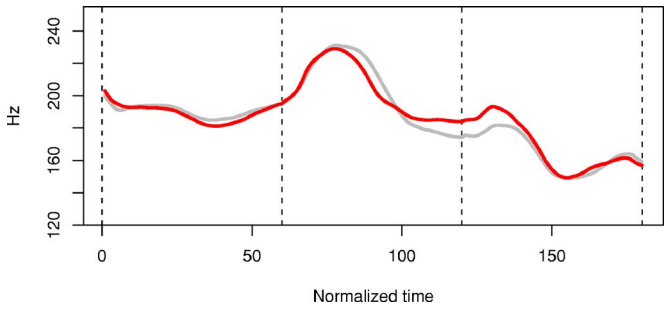


Fig. 3. (Color online) $f_0$ contours averaged over both word order conditions in the broad focus condition. Word boundaries are depicted by vertical dashed lines. Each word has been time normalized and averaged separately using 60 equidistant time points. The compensation described in Sec. IV can be seen in the marked word order contour (light gray line).

 

relative prominences are relevant; in our case the whole adverbial phrase. The results are, moreover, in accordance with a somewhat similar study by Xu and Xu on the realization of focus in English.[8]

Second, the results showed an effect of syntactic structure with respect to both intensity and, most importantly, tonal shape. In other words, the word order of the produced utterances modulated the difference between the two peaks, that is, their relative prominence. Whereas in perception, marked word order modulates *ceteris paribus* the perceived relative prominence of the two peaks by resulting in a decrease of prominence of the first peak and an increase of the second peak compared with the neutral unmarked case, in production speakers compensate, particularly in the broad focus condition, to keep the overall pattern neutral when the word order in fact signals a narrow focus on the later word. That is, they compensate to keep the whole utterance under sentence focus. The speakers clearly take into account the fact that the marked word order itself already signals focus, and compensate by enlarging the difference between the two peaks in the broad focus condition, in order to not end up unintentionally signaling a narrow focus on either word.

Whether the observed relation between syntax and prosody in focus assignment is specific to Finnish only remains for further investigation to find out. However, as Donati and Nespor[9] argue, prosodic focus and syntactic structure tend to be related, in that the more prosodic prominence is allowed to move around in the phrase the more rigid the word order properties of the language tend to be. Thus, it may be that the interaction between syntax and prosody in focus assignment is at its clearest with languages such as Finnish, where the tradeoff between intonation and syntactic structure for focus placement is sufficiently large due to the well-defined pragmatic functions of the word order changes. Accordingly, the interaction between syntactic structure and prosody in focus assignment may be less pronounced both in languages with either syntactically more constrained word order or syntactically free but pragmatically less-constrained word order.[4] Be that as it may, the present study suggests that, as in the perception of prosodic prominence, higher order structural information, word order, interacts with the basic prosodic parameters in order to ensure a semantically and pragmatically coherent interpretation of the utterance.

## Acknowledgments

Both authors have contributed equally to this paper. We would like to thank Mietta Lennes, Leena Wahlberg, and Anna Dannenberg for their help with obtaining and labeling the data for this study. We would also like to thank Stefan Werner for providing much help at earlier stages of the study reported here. We would also like to thank Daniel Hirst for his insightful comments on an earlier version of the manuscript. The present study was supported by Grant No. 107606 from the Academy of Finland to M. Vainio and Grant No. 106418 from the Academy of Finland to J. Järvikivi.

## References and links

[1]M. Vilkuna, *Free Word Order in Finnish: Its Syntax and Discourse Functions* (Suomalaisen Kirjallisuuden Seura, Helsinki, 1989).
[2]J. Pierrehumbert, "The perception of fundamental frequency declination," J. Acoust. Soc. Am. **66**, 363–369 (1979).
[3]C. Gussenhoven, B. H. Repp, A. Rietveld, H. H. Rump, and J. Terken, "The perceptual prominence of fundamental frequency peaks," J. Acoust. Soc. Am. **102**, 3009–3022 (1997).
[4]M. Vainio and J. Järvikivi, "Tonal features, intensity, and word order in the perception of prominence," J. Phonetics **34**, 319–342 (2006).
[5]C. Shih, "A declination model of Mandarin Chinese," in *Intonation: Analysis, Modelling and Technology* (Kluwer Academic, Dordercht, 2000), pp. 243–268.
[6]P. Boersma, "PRAAT, a system for doing phonetics by computer," Glot International **10**, 341–345 (2001).
[7]F. E. Harrell, *Regression Modeling Strategies* (Springer, New York, 2001).
[8]Y. Xu and C. X. Xu, "Phonetic realization of focus in english declarative intonation," J. Phonetics **33**, 159–197 (2005).
[9]C. Donati and M. Nespor, "From focus to syntax," Lingua **113**, 1119–1142 (2003).

  