Linguistic Bias, Tonal Features, and Intensity in the Perception of Prominence

Martti Vainio

Department of Phonetics, Department of General Linguistics, University of Helsinki, Finland

Juhani Järvikivi

Department of Psychology, University of Turku, Finland

Running title: Perception of Prominence
Abstract

Four experiments investigated the perception of prominence as a function of sentence stress in Finnish. Listeners judged the relative prominence of two consecutive nouns in a three-word utterance where the accentuation of the nouns was systematically varied by tonal means. Experiments 1 and 2 investigated both the tonal features underlying the subjects’ responses as well as the influence word order on the perceived prominence of the two accented words. The results showed that similar tonal features regardless of other phonetic and syntactic differences conditioned the subjects’ judgments of prominence. They further showed that changing the word order influenced the distribution of responses in the two experiments. Two further experiments were administered to check the possible influence of slight tonal and loudness differences in the first two experiments. Only loudness was found to affect the distribution of judgments. Furthermore, the influence was local and only affected the latter word. Overall, the results suggest that the most important tonal features responsible for the perception of prominence form a so-called flat hat pattern. They also indicate that different kinds of focus structure influence the perception of prominence even when the judgments are based on decisions about the place of sentence stress.

Key words: intonation, focus, prominence, sentence stress, accentuation, Finnish, experimental phonetics

* Corresponding author.

Email address: martti.vainio@helsinki.fi, juhani.jarvikivi@utu.fi

(Martti Vainio).

1 Present address: P.O. Box 9 (Siltavuorenpenger 20), FIN-00014, University of Helsinki, Finland
1 Introduction

In the field of phonetics it is well-established that linguistic knowledge – mainly in the form of phonological systems – influences phonetic perception in a top-down manner. This is perhaps best seen in the phonemic perception of assimilated segments as their representative phonemes rather than as the actual phones they are realized. In a similar vein, Pierrehumbert (1979) found that although instructed specifically to pay attention to pitch, her subjects’ judgments were actually based on various other factors besides pitch alone. Moreover, studies in second language learning have shown that native language sounds are perceived more easily than those acquired later in life through a second language (see Hume and Johnson (2003) and references therein). That is, listeners interpret similar phonetic structures and units differently depending on their linguistic knowledge.

It can, therefore, be hypothesized that such perceptual influence should also be found within the less discrete linguistic and phonetic phenomena, such as prominence. Indeed, Eriksson, Thunberg, and Traunmüller (2001) have found such an effect concerning syllable prominence; in their study linguistically motivated factors explained the prominence ratings of syllables better than signal-based cues (57% and 48%, respectively). The influence of linguistic categories on phonetic perception has consequences for any study on the perception of (prosodic or syllabic) prominence. The situation becomes crucial when a seemingly similar prosodic structure can be present with distinctly different syntactic structures. This is the case, for instance, in Finnish where the syntactically (relatively) free word order can be used for pragmatic purposes, for example, to bring a given constituent in an utterance into focus without
changing the prosodic structure in any way. Therefore, we may expect the
prominence pattern in a sentence with unmarked word order, such as “Me-
nemme laivalla Lemille” (We go by boat to Lemi) to be perceived differently
with respect to prominence depending on the order of the two adverbs, laivalla
and lemille, in the adverbial phrase “laivalla Lemille” (“by boat to Lemi” as
opposed to Lemille laivalla “to Lemi by boat”). In other words, changing the
word order to “Lemille laivalla”, with an emphatic or contrastive focus on the
latter word, should be reflected on how prominent it is perceived to be.

We conducted a series of four experiments to study the perception of prom-
minence in a two-accent utterance in Finnish. We were interested in what are
the characteristic tonal factors in the intonation contour that modulate the
perception of prominence and whether they remain the same regardless of
different information structures represented by different word order permuta-
tions. More importantly, we were interested in whether word order by itself
would have an influence on the perception of prominence in such utterances.

1.1 Prominence and Categorical Perception

It is well-known that speakers can vary the prominence of pitch accents by
varying the height associated fundamental frequency ($f_0$) maxima to express
different degrees of emphasis (Gussenhoven, Repp, Rietveld, Rump and Ter-
ken (1997) p. 3009). Listeners react to these changes accordingly. That is,
the perceived prominence of any accented syllable is related to the height of
the fundamental frequency maximum as well as to the relation of that local
maximum to other maxima in the utterance. For instance, it has been shown
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 (see for instance Pierrehumbert (1979) for English, Gussenhoven et al. (1997) for Dutch and Vainio, Mixdorff, and Järvikivi (2003) for Finnish).

Terken and Hermes (2000) noted that we currently lack the sufficient knowledge to determine whether the perception of accent strength varies in a gradient way or not, although results from many experiments seem to support the assumption that the perception of prominence is, in fact, gradual. But if we view prominence as (partially) reflecting a linguistic category — such as focus — rather than as a gradually varying phonetic phenomenon, we must assume that the perception then becomes more or less categorical. The situation is much the same as with, say, formants, which give rise to categorical perception if we study them through vowel categories but are gradual if we study them directly. Thus, if we consider focus to be a discrete linguistic phenomenon, we must assume that the perception of focus must be categorical in a sense that it must divide the perceptual space at some point. Therefore, we can also assume that, as a linguistic category, focus must influence the perception of prominence in much the same way that phonemes (or different combinations of phonological features) influence the perception of segmental phonetic variables.

The categorical nature of intonation has been studied relatively little but some evidence for certain intonational phenomena being categorically perceived has been found. However, the evidence seems somewhat conflicting. Remijsen and van Heuven (2003) found evidence for categorical perception between Dutch boundary tones signaling statements and questions. In contrast, Ladd and Morton (1997) did not find such evidence for “normal” and “emphatic” accent peaks in English. Although it is not the purpose of the present study
to investigate categorical perception per se, we will inquire into whether the
given categorical boundaries measured from gradual prosodic variables may
vary depending on the linguistic or information structure of the utterance.

1.2 Focus: Word order and prominence

Apart from grammatical relations proper, the relative order of constituents
within a sentence as well as its phonology can be used to convey aspects
of the distribution of information within a sentence. This distribution of in-
formation is referred to as the information structure. An important part of
information structure has to do with the role of new - given - and old - in-
ferred - information. Although the terminology varies considerably, the given
or presupposed information is traditionally referred as the topic of the sen-
tence. In contrast, focus is usually used to refer to what is new, or, what is
not within what is pragmatically presupposed (e.g., Van Valin and La Polla,
1997). Many times, however, it is not just the relation between old and new
information that is important, but the relationship between them. In Finnish,
for example, the syntactically free word order is harnessed to serve informa-
tion structure. Thus, in an unmarked case, such as “menimme laivalla Lemille”
(we went by boat to Lemi), the canonical order of the two adverbs (manner +
place) in the adverbial phrase conforms to its information structure, and the
phrase as a whole can be said to be under so-called sentence focus (Van Valin
and La Polla, 1997). Consequently, no pragmatic presuppositions are evoked
by the word order. In contrast, however, changing the word order to marked
“menimme Lemille laivalla” presupposes the information that we did in fact
go to Lemi, but in this case, the word order is used to emphasize or focus the
fact that it was by boat we went to Lemi - and not by a car - as an answer, for instance, to a question “how did you go to Lemi?” (for the pragmatic use of word order in Finnish, see, e.g., Hakulinen and Karlsson (1979), and Vilkuna (1989)). Apart from word order, there is another means generally available for placing any of the constituents under the domain of focus even in the un-marked case, namely prosody. Prosodically focus can be achieved, for example, by increasing the accent or stress on the part of an utterance that is intended to be brought into focus. For instance Finnish can allow for any constituent in an utterance to be focused by prosodic means: thus a Finnish speaker can say “Manne meni Lemille” ("Manne went to Lemi") as well as “Manne meni Lemille” ("Manne went to Lemi”; italics depict focus). Thus, it is of interest how the two main means available - syntactic and prosodic - for the marking of focus affect the perception of one or another part of an utterance as more or less prominent than the others.

In the present paper the influence of the accent strength and word order on perception of prominence was studied with a series of perception experiments. The experiments described here fall in line with a series of somewhat similar studies reported by e.g. Pierrehumbert (1979), Gussenhoven and Rietveld (1988), Terken (1994) and, Ladd, Verhoeven, and Jacobs (1994), as well as Gussenhoven et al. (1997), which deal with the perception of prominence in an utterance with two accented words in the form of two $f_0$ peaks on the accented syllables. In this paper we use the term prominence to refer to the auditory salience of a phonetic or a linguistic unit. With sentence stress we refer to the utterance level prominence relations between words. In the framework of our study, sentence stress can be seen to signal emphatic or contrastive focus. The perception of prominence has generally been studied in relation to tonal
features and their dynamics (see, for instance Terken (1989), Terken (1994), Gussenhoven et al. (1997), Hermes (1997), and Terken and Hermes (2000)). Most of the studies listed above attempt to relate the $f_0$ variation to perceived prominence in order to develop a metric for prominence (Gussenhoven et al., 1997). All of the earlier studies make clear that listeners estimate the prominence of the pitch peak on the basis of the pitch characteristics of the contour around it (Gussenhoven et al., 1997). However, none of them explicitly examine the possibility that syntax and information structure may influence prosodic perception. In fact, some of the studies use delexicalized utterances and, thus, avoid the problem. Although, this probably does not have consequences with regard to the published results, it may have consequences with regard to their explanations. In other words, they do not take into account the possibility that there may be other than signal-based factors which influence listeners’ prominence estimates. The main difference between the present study and the ones listed above is that all of the previous studies concentrated on prominence as a phonetic phenomenon, whereas in the present study we were interested in both the tonal means as well as word order with only indirect implications to prominence as it is realized through sentence stress or accent alone (depending on the terminology in use). The role of other prosodic parameters – mainly intensity and segmental durations – in the perception of prominence has also been investigated, but not as systematically and to a much lesser degree. Especially the relative intensity within an utterance and its influence on the perception of prominence has not been systematically studied to the best of our knowledge (excluding, of course, Pierrehumbert’s 1979 paper, where she studied intensity (amplitude) in one of the experiments). This is regrettable, especially since Batliner, Buckow, Huber, Warnke, Nöth, and Niemann (2001) have shown that duration and energy features are more important than $f_0$ for
both English and German accent classification. How much of this holds for Finnish, is to be determined. However, it has interesting consequences as to the results from the current Experiment 2 (see Section 2.2 for more detail). The next section describes the present series of experiments in more detail.

2 Experiments

Experiments 1 and 2 were conducted in order to investigate the perception of prominence in Finnish. More precisely, we investigated the tonal features responsible for prominence as well as the influence of syntactic structure, namely, word order, on the perception of prominence.

In Experiment 1 the sentence “Menemme laivalla Lemille” (We go to Lemi by boat) was used. The sentence permits four possible interpretations with respect to the location of focus:

(1) broad (or sentence) focus: Menemme laivalla Lemille. (e.g., as in an answer to “What will you do tomorrow?”)

(2) narrow focus on “laivalla”: Menemme laivalla Lemille. (e.g., “by boat” as an answer to “How are you going to Lemi”)

(3) narrow focus on “Lemille”: Menemme laivalla Lemille. (e.g., “to Lemi” as an answer to “Are you going to Luumäki [a place near Lemi] with a boat?”)

(4) multiple contrastive, narrow focus on both “laivalla” and “Lemille”: Menemme laivalla Lemille (e.g., as an answer to “So you are going to Luumäki by train?”)

Illustrative $f_0$ contours for each interpretation are shown in Figure 1. Only the
first three conditions were investigated in the current study.

*** Insert Figure 1 about here ***

For each experiment a set of 125 stimulus utterances was constructed: the baseline declination was set at five different levels and also the accentuation of the two nouns was varied in five levels. Thus, the stimuli covered a $5 \times 5 \times 5$ array ranging from complete de-accentuation to emphatic accent on each of the nouns with a varying baseline declination. A schematic representation of the stimulus parameters can be seen in Figure 3 in Section 2.1.

*** Insert Figure 2 about here ***

We employed the Fujisaki model (Fujisaki and Hirose, 1984) as a means to produce phonetically constrained stimuli for the experiments. The rationale for using an intonation model which produces smoothly varying contours was based on the fact that such a model captures the underlying form of the $f_0$ curve in a fairly reliable manner; i.e., a smooth contour is free of so-called microprosodic variation, which is generally considered to be segmentally conditioned, and, therefore, irrelevant to the research questions at hand. The model also allows for an effective way to manipulate a set of stimuli by varying the model parameters directly. The model parameter values for the accent amplitude values and their corresponding values in semitones and Hertz are listed in Table 1. The model has been previously applied to Finnish with success (Mixdorff et al. (2002) and Vainio, Mixdorff, and Järvikivi (2003)).

*** Insert Table 1 about here ***
2.1 Experiment 1

2.1.1 Participants

Twelve phonetically untrained students from a linguistic graduate summer school at the University of Helsinki participated in the experiment. All were native Finnish speakers and none reported any hearing loss. None of the participants were involved in speech research.

2.1.2 Materials

The utterance “Menemme laivalla Lemille” with two pitch accents, on the first syllable of both “laivalla” and “Lemille” was chosen as a starting point for building a set of prosodically manipulated stimuli. The baseline stimulus for the experiment was chosen from a set of utterances produced to elicit a broad focus condition in the adverbial phrase for another study on the suitability of the Fujisaki model for Finnish (Mixdorff et al., 2002). In that study, a group of subjects determined the naturalness of the utterances and the suitability of the utterances’ membership in the intended sentence category. The utterance which was judged to be the most natural with regard to the mean opinion score was chosen as the baseline stimulus for the present experiment. Furthermore, it was checked that the selected utterance received unanimous judgment as to belonging to the intended category of broad focus; no single word was perceived to be more prominent than others.

The original utterance from the Mixdorff et al. (2002) study used as a basis for stimulus construction here was originally recorded in an anechoic chamber at the Acoustics Laboratory at the Helsinki University of Technology with
a high-quality microphone and pre-amplifier and quantized at 16 bit at 41.1 kHz sampling frequency. The utterance was then analyzed with a robust pitch detection algorithm, and the Fujisaki model parameters were estimated by automatic means and manually corrected to fit the original $f_0$ track. This yielded an initial set of parameter values which were then varied systematically to produce the intended $f_0$ contours for the present stimuli (see Figure 3 for more detail).

In order to avoid any voiceless gaps, the utterance was designed to consist of voiced segments only. The first accent peak rise started at 0.49 seconds and peaked at 0.72 seconds; the second peak rise started at 0.99 seconds and peaked at 1.23 seconds. There was always a fall between the peaks. The utterance ended with a vocal fry during the last two syllables of the utterance.

*** Figure 3 about here ***

A set of parametrized pitch contours was produced by systematically varying the Fujisaki model phrase and accent components to produce a continuum of stimuli in a three dimensional “accent space”, as depicted in Figure 2. The resulting pitch contours were superimposed on the original utterance with a time domain PSOLA (pitch-synchronous overlap and add) method. This procedure left all other prosodic cues (except the microprosodic variation) intact.

2.1.3 Procedure

The stimuli were randomized and presented to the subjects through high-quality headphones at comfortable loudness levels in a quiet class-room in
blocks of 26 stimuli. Each block was preceded by one second long sine tone of around 400 Hz, and a fifteen second pause was inserted between each block. The inter-stimulus interval was set to 4 seconds. Ten practice trials preceded the first experimental trial.

The participants were instructed to indicate on an experimental sheet whether they perceived the main stress of the utterance to be: (1) on the word “Lemille”, (2) on the word “laivalla” or (3) on neither of them. The intuitively clear concept of stress was meant to encourage the subjects to be in a more linguistic, and, therefore, more synthetic listening mode, which is considered more suitable for prominence judgments in general (Gussenhoven et al., 1997). Moreover, Pierrehumbert 1979 suggested, that regardless of instructions to specifically pay attention to pitch, the subjects in fact, made judgements based on relative prominence which is the result of various factors besides pitch alone.

2.1.4 Results and Discussion

Figure 2 shows the basic $f_0$ contour used in the experiments and lists the points of interest ($t_1 - t_6$) and the dynamic factors ($t_i - t_j$) used in subsequent statistical analyses. For instance, the factor $t_2 - t_3$ stands for the amount of rise of the first peak. The effect of each operationalized factor was determined by regression analysis. We further assessed the effects of the baseline declination and the peak sizes on the subjects’ responses with analyses of variance

\footnote{The baseline stimulus with the parameter values of the original utterance was inserted to each stimulus block in order to check any possible effects caused by hearing the same utterance multiple times.}
(ANOVA)s.

The responses for the different sentence stress conditions were pooled by items and multiple regression analyses were conducted to determine the factors which best predicted the variance of the responses. The results were analyzed separately for response categories 1 and 2 excluding the negative category 3 (stress perceived on neither of the two words). The results from the regression analyses showed that the only factors explaining the variance of Response 1 (sentence stress perceived to be on the first noun; “laivalla”) were the difference between the two peak heights; $t_3 - t_5$ in Figure 2 ($t = 12.30, p < .001$) and the amount of rise in the first peak; $t_2 - t_3$ in Figure 2 ($t = -5.24, p < .001$). The overall regression model with the above two factors was highly significant; $F(2, 127) = 182.0, R^2 = 0.74, p < .001$. In other words, the most important factors modulating the perception of prominence were found to be the difference between the points $t_2$ and $t_3$ – the magnitude of the rise of the first peak – as well as the difference between points $t_3$ and $t_5$, i.e., difference between the two peak maxima.

The results for Response 2 (sentence stress perceived to be on the second noun (“Lemille”)) were in turn explained by a model, which included, again, the difference between the two peaks ($t = -11.05, p < .001$) and the fall of the latter peak, i.e., the difference between points $t_5$ and $t_6$ ($t = 6.51, p < .001$). Again, the overall regression model was highly significant $F(2, 127) = 191.9, R^2 = 0.75, p < .001$. In this case, the most important factors were the difference in peak maxima and the magnitude of the fall of the last peak.

Thus, the two peaks, which have a superficially similar tonal structure, turned out to be different from a perceptual point of view. Although, the most im-
portant feature modulating the perception of prominence for both *Response 1* and *Response 2* was the difference between the peak maxima, it was the fall rather than the rise which affected the results of *Response 2*. This is different from what modulated *Response 1* and what could also be expected intuitively, that is, the magnitude of rise.

Moreover, as suggested by Figure 4, the stress was perceived unanimously to be on the latter word, when the first peak was lower than the second one in absolute terms. We will return to this in general discussion.

To assess the effects of the manipulation of the phrase component (leading to changes in baseline declination, hence Phrase) and accent components (leading to changes in peak heights, hence Accent) on the participants’ responses, ANOVAs were carried out on the participant means with Phrase (five levels) and Accent (five levels) as within-participant factors. The analyses were done separately for the *Response 1* and *Response 2* categories with Accent standing for the first and second pitch peak, respectively.

*Response 1*. ANOVAs revealed a significant main effect of Phrase \([F(4, 44) = 8.61, p < .001]\) as well as Accent \([F(4, 44) = 107.24, p < .001]\). There was also a significant interaction of Phrase and Accent \([F(16, 176) = 1.92, p < .05]\).

*Response 2*. There was again a significant main effect of Phrase \([F(4, 44) = 9.26, p < .001]\) as well as Accent \([F(4, 44) = 76.03, p < .001]\). There was also a significant interaction of Phrase and Accent \([F(16, 176) = 2.71, p < .01]\).

The main effects of Accent in both response categories are obvious in the light of the regression analyzes and do not call for further elaboration. The rate of the baseline declination (Phrase) was not of primary concern for us, but we
did expect it to have an effect on the responses based on previous research (Gussenhoven et al., 1997), which states that with respect to *Response 2* the second peak needs to be higher in stimuli with steeper declination in order to gain prominence. The observed effect will be discussed further in combination with Experiment 2 in Section 2.2.4. The phrase effect on *Response 1* is, again, obvious from the fact that the declination directly affects the first peak height.

As to the interactions, they are relatively weak and occur at the extremities of the stimulus space. The interactions were then caused by changes in the perception of prominence within the set of stimuli that can be regarded as somewhat unnatural – especially, when considering the fact that concomitant changes in other prosodic parameters were not manifested in the stimuli. That is, an extremely high $f_0$ peak usually co-occurs with severely lengthened segmental durations, increased intensity, and changes in voice quality.

As the stimuli were designed to cover the accent space evenly, and also unmarked word order was used, we expected the responses to be evenly distributed between the different stress conditions. As it turned out, each category did in fact receive approximately a third of the responses (see Section 2.2.4 and Figure 5). However, if the perception of prominence is connected to other than purely phonetic factors, such as focus, we can expect to have a different distribution of responses if we focus by non-phonetic or non-prosodic means. As was mentioned above, in Finnish also word order can be used for such purposes. Thus, Experiment 2 investigates whether placing the focus on the latter word by changing the order of the two nouns in the adverbial phrase affects the prominence judgements as compared to Experiment 1. In other words, whether non-phonetic factors, here word order, have consequences as to how prominent the latter peak is perceived to be.
2.2 Experiment 2

The second experiment was similar to Experiment 1 in all other respects except that the word order of the adverbial phrase "laivalla Lemille" was switched to "Lemille laivalla". As in the case of "laivalla Lemille" the order of the two adverbials conforms to the canonical order of adverbs in adverbial phrase in Finnish, i.e., manner + place (e.g. Hakulinen and Karlsson, 1979), the word order manipulation resulted in the adverb of manner being in the second case moved to the emphatic or contrastive focus position. Thus, as an answer to a question, "with what are you going to Lemi", the phrase "menemme Lemille laivalla" would translate into English as "it is by boat we are going to Lemi". The question now becomes whether the word order manipulation would have consequences as to the perception of the relative prominence of the two adverbial NPs. If perception of prominence is influenced by factors other than the pitch alone, as argued by Pierrehumbert (1979) and Eriksson et al. (2001), we expect also the word order manipulation to affect the subjects' prominence judgements. More precisely, we expect the prominence to be perceived more often on the second peak as compared to Experiment 1. However, this should not affect the overall underlying tonal features responsible for the perception of prominence, and therefore we expect to observe a similar pattern of tonal factors for both peaks as was observed in Experiment 1.

2.2.1 Participants

Ten students from the Department of Linguistics at the University of Helsinki participated in the experiment. All were native Finnish speakers and none reported any hearing loss. None of the participants were involved in speech
2.2.2 Materials

The sentence “Menemme Lemille laivalla” (“We go to Lemi by boat”) was recorded by the same speaker as in Experiment 1 in noise-free room using a high-quality microphone placed approximately 5 cm from the speaker’s mouth. The recording was done directly to a computer using a professional level analogue-to-digital transformer. The recording was done with 16 bit quantization using 44.1 kHz sampling frequency. The utterance was then subjected to similar procedure as described in the previous experiment, and a similar set of 125 materials was produced. Since the original stimulus was tonally very similar to the one in Experiment 1, only the timing of the accentuation parameters had to be changed. This had certain consequences with regard to the stimuli (see Section 2.1.4 for more detail). The resulting stimuli were, thus, nearly identical to the ones in Experiment 1.

2.2.3 Procedure

The experimental procedure was identical to that in Experiment 1.

2.2.4 Results and Discussion

As with Experiment 1, the responses for the different sentence stress conditions were pooled and multiple regression analyses were conducted to determine the tonal factors which best predicted the variation within the responses. The results show that the only factors explaining the variance of Response 1 (sentence stress perceived to be on the first noun; “Lemille”) were the dif-
ference between the two peak heights; $t_3 - t_5$ ($t = 11.33, p < .001$) and the amount of rise in the first peak; $t_2 - t_3$ ($t = 3.76, p < .001$). The overall regression model was highly significant ($F(2,127) = 131.6$ and $R^2 = 0.67$, $p < .001$). The results for Response 2 (sentence stress perceived to be on the second noun (“laivalla”)) were in turn explained by a model, which included, again, the difference between the two peaks ($t = -10.79, p < .001$) and the fall of the latter peak ($t = 8.37, p = .001$). Again, the overall regression model was highly significant [$F(2,127) = 241.1$ and $R^2 = 0.79$, $p < .001$]. Unlike in Experiment 1 the rise of the second peak was marginally significant with regard to Response 2 ($t = 1.94, p = 0.0544$).

As in Experiment 1, the two $f_0$ peaks, have a similar structure from a perceptual point of view: the most important feature for the first peak was its rise while the latter peak was characterized by a fall. With regard to the actual contrast between the peaks, the difference between the peaks was, again, by far the most important factor. Moreover, the sentence stress was, again, unanimously perceived to be on the latter word when the first peak was lower than the second one in absolute terms (see Figure 4). However, the phenomenon was not as pronounced as in Experiment 1.

*** Insert Figure 4 about here ***

With respect to the effects of the phrase and accent components the results are as follows: Response 1; ANOVAs revealed a significant main effect of Phrase [$F(4,36) = 10.08, p < .001$] as well as Accent [$F(4,36) = 47.37, p < .001$]. There was also a significant interaction of Phrase and Accent [$F(16,144) = 1.99, p < .05$]. As to Response 2, there was again a significant main effect of Accent [$F(4,36) = 137.99, p < .001$]. However, there was no main effect of
Phrase \( F(4, 36) = 2.20, p > .08 \). There was also a significant interaction of Phrase and Accent \( F(16, 144) = 1.98, p < .05 \).

What is of interest here is the absence of the effect of Phrase for Response 2. This absence may have been due to either the changed word order or possible phonetic differences between the stimuli in the experiments or both. We will return to this in more detail in Section 2.3.4. Although it is likely that the manipulation of word order, thus the shift of focus to the second NP, affected the prominence judgments in such a way that the phrase component ceased to have an independent effect on peak two, it is possible that a small difference in the relative intensity of the peak two as compared to Experiment 1 also had an effect.

As in Experiment 1, the interactions are relatively weak and occur at the extremities of the stimulus space and were, again, caused by changes in the perception of prominence within the set of stimuli that can be regarded as somewhat unnatural due to the lack of accompanying changes in loudness and segmental durations.

Figure 5 shows the distribution of responses with regard to the sentence stress condition in Experiments 1 and 2. As was expected, the responses were distributed differently in the two experiments. Whereas in Experiment 1 the responses were evenly distributed between the three conditions, there were more responses for the second condition (stress on the latter word) in Experiment 2 and fewer responses for the first condition. The change in the distribution of judgements observed in Experiment 2 suggests that prominence was indeed affected by other than the purely tonal factors, such as pitch. In other words, it seems that the manipulation of the syntactic (and information) structure via
placing the focus on the latter word, attracted also more prominence judgments on the latter peak, despite the fact that the subjects were explicitly instructed to decide whether the stress in each sentence was on either word or not. More importantly, the placing of focus on the latter peak seems to have affected the prominence judgments of the first peak to a similar degree, although to an opposite direction.

*** Insert Figure 5 about here ***

It should be noted, however, that there were unavoidable, although small phonetic differences between the two sets of stimuli which could have influenced the distributions. The phonetic differences were mainly due to the origin of the baseline stimulus in Experiment 1. The experiment in Mixdorff et al. (2002) was not concerned with the relative prominence of the two accent peaks and the stimulus was not therefore designed to be symmetrical in that sense. It was chosen as a starting point for the current experiments on the basis that it was judged as the most natural token for broad focus from a set of seven repetitions of the utterance. There can be no such procedure for choosing a candidate for the other word order condition on the basis that, while producing a neutral candidate with regard to the focusing function of the word order change, the speaker might, in fact, compensate negatively for the prosodic prominence. That is, the second word order condition is not neutral with regard to focus to begin with, and, thus, it could also influence the relative prominence of the constituents. Therefore, the basic stimulus for the experiment was chosen from a set of utterances, where the speaker clearly intended to produce prosodically as neutral or broad a focus as possible. On post hoc analysis, however, this turned out to have affected some of the phonetic characteristics of the utterance. First, the relative intensity between the two accent peaks was different.
from the utterance in Experiment 1 (approximately −5 dB as opposed to approximately +1 dB). Second, the different segmental make-up of the accented nouns caused the $f_0$ contours to be slightly different after the second peak – i.e., the degree of fall was slightly larger (approximately 2 semitones on the average) in the second set of stimuli.

In order to counter for the slight possibility that the observed difference in Experiment 1 and Experiment 2 was affected by these factors we designed two further experiments to investigate the role of the degree of fall in the latter accent peak (Experiment 3) and the role of loudness (Experiment 4). For considerations of space, Experiments 3 and 4 will be discussed together.

2.3 Experiments 3 and 4

Experiment 3 was designed to investigate the effect of the small tonal differences between the stimuli in the first two experiments, whereas Experiment 4 was designed to investigate the perceptual effect of the observed loudness differences in the stimuli between Experiments 1 and 2.

One phonetic difference discussed in relation to the second experiment was the greater degree of fall in the last $f_0$ peak in the stimuli of the second as opposed to the first experiment. Since the fall was shown to be mainly responsible for the perception of its prominence, it is possible that the different results between the two experiments were at least partly caused by this systematic difference in the stimuli of the two experiments. If this was indeed the case, we expect the responses in Experiment 3 to be distributed similarly with the second experiment. If, however, the small difference in the degree of the fall
had no effect, we expect to end up with results distributed similarly with the first experiment.

As noted above, also the relative loudness between the two accented words in the two experiments was different in that the latter noun in Experiment 2 was considerably louder. Experiment 4 was therefore a replication of Experiment 2 in all other respects except that the intensity difference was controlled to correspond to the difference in Experiment 1. The observed intensity differences as well as the corrected intensity contour can be seen in Figure 6.

*** Insert Figure 6 about here ***

As intensity is usually considered to have an effect on the perception of prominence, we wanted to check whether the observed differences in the responses between the first two experiments were, in fact, due to the systematic intensity difference between the two sets of stimuli, and not the difference in word order. If that was indeed the case, the perceptual boost gained from the added intensity in the second experiment would be dampened in Experiment 4, and we could expect the responses in this experiment to be distributed more or less like in the first experiment.

2.3.1 Participants

Twelve and fourteen students from the Department of Linguistics at the University of Helsinki participated in Experiments 3 and 4, respectively. All were native Finnish speakers and none reported any hearing loss. None of the participants were involved in speech related research and none had participated in Experiments 3 and 4.
2.3.2 Materials

**Experiment 3.** The stimuli for the experiment were constructed by using the baseline stimulus from Experiment 1 and superimposing the pitch contours with a larger degree of fall after the second peak. The fall in Experiment 1 was, on the average 4.26 semitones as opposed to 6.22 semitones in Experiment 2. The pitch contours were otherwise similar to the ones in Experiment 1.

**Experiment 4.** The stimuli from Experiment 2 were multiplied with a simple linearly interpolated intensity contour in such a way that the relation of the average intensity of the two nouns resembled that of the stimuli in Experiment 1. Everything else in the stimuli was left intact.

2.3.3 Procedure

The procedure was identical to the previous experiments.

2.3.4 Results and Discussion

**Experiment 3:** The effect of the degree of fall. Two participants were discarded due to their having mainly responded with category 3. Our main concern was whether the responses would be distributed in the same manner as in Experiment 1 in which case we could determine that the tonal differences between the stimuli in Experiments 1 and 2 were not responsible for the distributional differences. Indeed, the responses were distributed almost exactly as in Experiment 1. A $\chi^2$ test on the proportions shows the difference between the results is non-significant ($\chi^2(2) = 4.85, p = 0.78$). This clearly shows that the greater degree of fall could not have caused the different results in the
first two experiments. The proportions of responses for all experiments are summarized in Table 2.

*** Insert Table 2 about here ***

**Experiment 4:** The effect of intensity. Figure 7 shows the distribution of responses in Experiment 4 against the responses in Experiment 1 and Experiment 2. It is immediately evident that lowering the intensity of the accented word leads to fewer responses in the second category, i.e., sentence stress on the latter word. What is remarkable, is that while the responses for the latter peak, where the change in intensity occurred, shifted to the third category, the relative number of responses in category 1 remained the same. Thus, the differences between the two experiments suggest that the effect of intensity is local to the latter peak and has no global effect.

Regression analyses for the responses of Experiment 3 and 4 revealed an identical pattern of tonal structures modulating the judgements of prosody as was found for Experiments 1 and 2.

*** Insert Figure 7 about here ***

As the only difference between Experiments 2 and 4 was the relative intensity between the latter of the two accented nouns, we analyzed the effect of loudness by pooling the responses of both experiments for both linear regression analyses and repeated measures ANOVAs. When added as a regressor, loudness was significant with respect to Response 2 ($t = 4.33, p < .001$) but not with respect to Response 1 ($t = -1.286, p = 0.20$). ANOVAs with Phrase and Accent as within-participants and Experiment (XP2 and XP4) as between-participants measures showed no difference between the experi-
ments on Response 1 ($F < 1$). Also the interactions between Experiment and the two other measures were non-significant. However, there was a significant difference between the experiments [$F(1, 22) = 6.20, p < .05$] as well as a significant interaction between Experiment and Accent in Response 2 [$F(4, 88) = 2.94, p < .05$]. This clearly shows, that the effect of loudness is local to the latter word or accent and has no global effect. Therefore the decreased number of responses for the first category in Experiments 2 and 4 can only be explained by the linguistic difference between the stimuli in the different experiments whereas the increased number of responses in the second category in Experiment 2 are mostly due to increased intensity of the latter word. The increased intensity can also explain the absence of the Phrase main effect for Response 2 in Experiment 2; that is, the intensity clearly decreases the importance of the tonal features with regard to prominence and, at the same time, it increases the local effect of the stressed syllable. Therefore, the syllable does not need to be raised tonally to compensate for a greater degree of baseline declination.

3 General Discussion

The present article reported the results from a series of four experiments investigating two distinct questions about the perception of prominence in Finnish: First, we investigated the tonal aspects and structures responsible for the perception of sentence stress in an utterance with potentially two stressed constituents (here single words). Second, we inquired into the possible influence of word order on the perception of relative prominence of the two constituents. In other words, we investigated whether focusing one of the
words by changing the word order would influence the subjects’ judgments of prominence. The first two of the reported experiments dealt with the tonal aspects of sentence stress and the influence of word order, respectively. The latter two experiments were designed to investigate two possible confounding factors between the first two experiments, namely dissimilarities in intensity levels and dissimilarities in tonal features in the respective stimulus materials.

The results showed a clear involvement of syntactic structure – word order - in the subjects’ perception of the relative prominence of the two accented peaks. More precisely, when the latter word was brought into focus with word order marking, it was also perceived as being stressed more often than the last noun in Experiment 1, where unmarked word order was used. As the possible confounding influences were ruled out in subsequent experiments, the results clearly show that, *ceteris paribus*, there is a clear top down effect of linguistic structure also on the perception of such phonetic phenomenon as stress.

With regard to the tonal structure of the utterances, the following main findings were obtained:

- First peak had to be lower than the second peak in order for the listeners to unanimously perceive the stress to be on the second peak.
- Both stress conditions were dependent on both absolute and relative difference between the two $f_0$ peaks.
- The prominence of the first peak was mainly dependent on the rise of the peak.
- The prominence of the second peak was mainly dependent on the fall of the peak.

What emerges from the results of all of experiments is the so-called *flat hat*
pattern which is generally characterized by a rise of the first pitch peak, a slope between the two peaks, and a fall of the last peak. Such a pattern has not been previously discussed in the literature on Finnish intonation and its existence in actual production is unclear. Generally, falling accents have not been attested in Finnish in a proper manner, but their existence is obvious when looked at informally: First, one can easily produce synthetic speech where a mere fall in an intonation contour can function as an accent. Second, falling accents can be found in frequently occurring lexicalized phrases, such as, the greeting “hyvää päivää” (good day) as opposed to a regular noun phrase “kolme hyvää päivää” (three good days), where there is a regular rise-fall pattern on the stressed syllable of the last word instead of just a fall as in the greeting case. The rise probably indicates a boundary, while the fall functions more as prominence lending. Nevertheless, the hat pattern occurs in all of the experiments pointing to a conclusion that the perception of prominence is determined by similar tonal structures regardless of other factors such as segmental durations and intensity. The actual degree of prominence is, however, determined by a combination of cues, not all of which are signal based.

Although not directly comparable, most of our findings are in agreement with similar findings from other languages. For instance, the fact that the latter peak in a two peak utterance has to be lower in absolute terms than the previous peak in order not to perceived as stressed, is directly comparable to the results reported for Dutch by Gussenhoven et al. (1997) and American English by Pierrehumbert (1979). The fact that the fall of the first peak and the rise of the second peak turned out not to be decisive to the prominence relations of the two peaks does not indicate that the sagging between the peaks is unimportant. It should be noted here that Finnish has a fixed stress
on the first syllable of the word and the rise associated with this syllable is a very important word boundary cue (Tuomainen et al., 1999; Tuomainen, 2001). Although we are inclined to argue that the boundaries signaled by rises in this way are those of phonological words rather than lexical ones. The “sagging transitions” are also in line with similar findings for English by Ladd and Schepman (2003).

As mentioned before, the role of intensity in the perception of prosody is very much an uncharted territory. Nevertheless, Pierrehumbert (1979) did study intensity (or amplitude) in a comparable setting to ours. According to her study the amplitude effect is 1.5 Hz/dB with regard to the so called crossover point where the two $f_0$ peaks are thought to be perceived as equally prominent. That is, the increased amplitude during the latter peak increases its prominence so that the crossover point is lower by 1.5 Hz for each increased dB in amplitude. In both Gussenhoven et al. (1997) and Pierrehumbert (1979) the crossover points for the latter peak in a two-peak utterance were below the maximum of the first peak in absolute terms. This is considered to indicate that listeners correct or normalize for the baseline declination. It is therefore interesting to see how the phonetic and linguistic factors studied here influenced the crossover points in the experiments.

We estimated the crossover points with probit analysis (Venables and Ripley, 1996) for both responses in our experiments by pooling the responses for all participants. According to Pierrehumbert’s theory the difference between the points in Experiment 1 and 2 should have been approximately 8 Hz (for an approximately 5 dB increase in intensity), which is exactly the value we obtained (the crossover points for experiments 1 and 2 being $-5.7Hz, SE = 0.7934$ and $-13.7Hz, SE = 0.8432$, respectively). Nevertheless, when we look at the res-
ults for Experiment 3, we do not get such a marked difference ($-8.3Hz, SE = 0.6728$). These results should be interpreted with utmost care. However, they do seem to suggest that the intensity difference cannot explain the observed differences for the Response 2 altogether. Furthermore, it is highly unlikely that the differences were due to tonal factors (i.e., the greater fall in Experiment 2), for the crossover point for Experiment 4 was only slightly different from Experiment 1 ($-5.2Hz, SE = 1.0279$).

4 Conclusion

We conclude that there is a clear and measurable linguistic bias in the perception of prosodic prominence in Finnish. We base this conclusion on the results from Experiment 1 and 2 as well as the fact that we ruled out all other factors but the different word order as an explanation for the differences in the judgments of prominence of the first NP in the utterances. Furthermore, phonetic differences could not explain all of the differences in the perception of prominence of the second NP either. All of the factors studied are in one way or another related to declination (top-line as well as baseline), which seems to, as in other languages, have a mental representation for Finnish listeners. This mental representation of declination is, however, more complex than was previously thought, for it refers not only to measurable, physical characteristics of the speech signal but abstract linguistic characteristics of the utterance as well.

Thus, the $f_0$ contour gives rise to the phonetic form, and consequently, the phonological distinctions related to accentuation, whereas the prominence relations in an utterance are determined by the magnitude of change in the pitch
excursions, the intensity of the syllables and the syntactic structure represented by word order.

Acknowledgements

We would like to thank Jukka Hyönä, Pirita Pyykkönen and Hanna Westerlund for their insightful comments on the manuscript. We also thank Stefan Werner for his contributions to the discussions on this subject.

References


Pierrehumbert, J. (1979). The perception of fundamental frequency declina-


Figure 1. The $f_0$ contours for the four possible focus conditions for a Finnish sentence “menemme laivalla Lemille” (we go by boat to Lemi).
Figure 2. A example $f_0$ contour with an orthographic transcription on syllabic basis. The points of interest in the contour are marked with $t_n$. The double-ended arrows depict the actual factors used for statistical analyses.
Table 1
Declination and peak values in Experiment 1. ap stands for the Fujisaki model phrase command amplitude and the aa1 and aa2 the accent command amplitudes for the peaks 1 and 2, respectively. The Hertz values for the peaks are means for the 25 different declination and peak conditions. Negative values are due to the baseline declination as the peak values are calculated from two distinct time values (beginning of rise and peak) in the $f_0$ contours.

<table>
<thead>
<tr>
<th>Declination</th>
<th>Peak 1</th>
<th>Peak 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ap</td>
<td>st/sec</td>
</tr>
<tr>
<td>1.2</td>
<td>7.7</td>
<td>0.46</td>
</tr>
<tr>
<td>1.0</td>
<td>6.4</td>
<td>0.34</td>
</tr>
<tr>
<td>0.8</td>
<td>5.0</td>
<td>0.23</td>
</tr>
<tr>
<td>0.6</td>
<td>3.6</td>
<td>0.11</td>
</tr>
<tr>
<td>0.4</td>
<td>2.9</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Figure 3. A three-dimensional view of the $125 f_0$ contours of the stimuli used in the experiments. The contours are ordered so that each declination type is cycled over, then the second peak and finally the first peak. The first stimulus is, thus, has a low declination value and no accents and the last stimulus has high declination as well as high accent values.
Figure 4. Distribution of second peak responses vs. the absolute difference between the two peaks for Experiments 1 and 2. The crossover points provided by the probit analyses are marked with a dotted line. The crossover points are described in Section 3.
Figure 5. Distribution of responses in Experiments 1 and 2. The squares depict Experiment 1 and the circles Experiment 2.
Figure 6. Intensity curves of the baseline stimuli in Experiment 2 (solid line) and Experiment 4 (dotted line).
Table 2
The percentages of responses for all four experiments.

<table>
<thead>
<tr>
<th>Response</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
<th>Experiment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response 1</td>
<td>32.3</td>
<td>22.0</td>
<td>33.4</td>
<td>24.3</td>
</tr>
<tr>
<td>Response 2</td>
<td>32.2</td>
<td>42.3</td>
<td>32.5</td>
<td>33.6</td>
</tr>
<tr>
<td>Response 3</td>
<td>35.5</td>
<td>35.7</td>
<td>34.1</td>
<td>42.1</td>
</tr>
</tbody>
</table>
Figure 7. Distribution of responses in Experiments 1, 2, and 4. The squares depict Experiment 1, circles Experiment 2, and diamonds Experiment 4.