Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Phonetics ▮ (▮▮▮▮) ▮▮▮–▮▮▮
Journal of Phonetics
www.elsevier.com/locate/phonetics

Tonal features, intensity, and word order in the perception of prominence ☆

Martti Vainio[a,b,*], Juhani Järvikivi[c]

[a]*Department of Speech Sciences, University of Helsinki, P.O. Box 9 (Siltavuorenpenger 20), FI-00014, Finland*
[b]*Department of General Linguistics, University of Helsinki, P.O. Box 9 (Siltavuorenpenger 20), FI-00014, Finland*
[c]*Department of Psychology, University of Turku, Assistentinkatu 7, FI-20014, Finland*

Received 8 April 2004; received in revised form 12 May 2005; accepted 24 June 2005

Abstract

The perception of prominence as a function of sentence stress in Finnish was investigated in four experiments. 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 of word order on the perceived prominence of the two accented words. The results showed that similar tonal features regardless of other phonetic 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 intensity differences in the first two experiments. Only intensity was found to affect the distribution of judgments. Furthermore, the influence was local and only affected the last of the two words. Overall the results suggest that the most important tonal features responsible for the perception of prominence form a so-called flat-hat pattern. That also indicates 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.

☆ Both authors have equally contributed to this paper.

*Corresponding author. Department of General Linguistics, University of Helsinki, P.O. Box 9, 00014 Helsingin yliopisto, Finland.
E-mail address: martti.vainio@helsinki.fi (M. Vainio).

0095-4470/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2005.06.004

1. Introduction

In the field of phonetics it is well-established that linguistic knowledge can sometimes influence phonetic perception in a top-down manner. This is perhaps best seen in phonemic perception where listeners recover canonical phonemes even when they are overlapped and possibly blended in an assimilation. Top-down processing has also been shown to work in phonemic restoration—a particularly powerful auditory illusion in which listeners "hear" parts of words that are not really there (Samuel, 1981). The underlying linguistic factors range from phonological to pragmatic. 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 & 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 a perceptual influence should also be found within 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 (linguistic factors, 57%; signal-based factors, 48%). The influence of linguistic categories on phonetic perception has consequences for any study of the perception of (prosodic or syllabic) prominence. The influence of top-down processing must be taken into consideration when a seemingly similar prosodic structure can be present with distinctly different syntactic structures. This is the case, for instance, in Finnish where 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 "Menemme 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 other words, changing the word order to "Lemille laivalla" (to Lemi by boat), with an emphatic or contrastive focus on the last word, should be reflected in how prominent the adverbs are perceived to be.

We conducted a series of four experiments to study the perception of prominence in a two-accent utterance in Finnish. We were interested in the characteristic tonal factors of 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 permutations. More importantly, we were interested in whether word order would have an influence on the perception of prominence in such utterances.

1.1. Prominence and discrete categories

It is well-known that speakers can vary the prominence of pitch accents by varying the height of the associated fundamental frequency (f_0) maxima to express different degrees of emphasis (Gussenhoven, Repp, Rietveld, Rump, & Terken, 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, & Järvikivi, 2003 for Finnish). Pierrehumbert (1979) explains this by postulating a mental representation of declination which is used by the listener to normalize for physically conditioned declination of f_0.

Terken and Hermes (2000) noted that we currently lack 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 may assume that the perception then becomes categorically interpretable. The situation is much the same as with, say, formants, which seem to give rise to categorical perception if we study stop place perception in CV syllables, but are gradual if we study them directly as acoustical entities (Blumstein & Stevens, 1980) or vowel perception in general (Winkler et al., 1999). Thus, if we consider focus to be a discrete linguistic phenomenon, we must assume that the perception of focus must be categorically interpretable in the 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, they too, found that the utterances were interpreted categorically. In any case, it is not the purpose of the present study to investigate categorical perception per se—which is itself a controversial issue—but rather to establish whether the given categorical interpretations with respect to the gradual prosodic variables are influenced by 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 information is referred to as the information structure. An important part of information structure has to do with the role of new (given) and old (inferred) information. Although the terminology varies considerably, the given or presupposed information is traditionally referred as the topic of the sentence. 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 & La Polla, 1997). Many times, however, it is not just whether the information status of a particular referent is "old" or "new" that is important. It is instead, often the relationship between a focused referent ("new" information) and what is pragmatically presupposed which together make the focused referent informative, not the fact that it is newly introduced. In Finnish, for example, the syntactically free word order can be manipulated to serve information 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 default information structure, and the phrase as a whole can be said to be under so-called sentence focus (Van Valin & La Polla, 1997) whose prosodic counterpart would be broad focus. 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 if it were an answer to a question "how did you go to Lemi?" (for the pragmatic use of word order in Finnish, see, e.g., Hakulinen & 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 unmarked case, namely prosody. Focus can be achieved prosodically by increasing the accent or stress on the part of an utterance that is intended to be brought into focus. In Finnish, any constituent can 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 prosodic 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 accent strength and word order on prominence perception 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. We use *sentence stress* to 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, 1994; Gussenhoven et al., 1997; Hermes, 1997; Terken & 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 the latter all concentrated on prominence as a phonetic phenomenon, whereas in the present study we were interested in how both the tonal means and word order give rise 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. In particular, the relative intensity within an utterance and its influence on the perception of prominence has not been as systematically studied as the tonal aspects of prosody. This is regrettable, especially since Batliner et al. (2001) have shown that duration and energy features are more important than f_0 for both English and German accent classification based on principal components analysis. There are, however a number of studies relating intensity and prosodic focus and prominence. In a production study, Heldner (1996) found an intensity difference between focused and non-focused words, which interacted with the position of the word in the sentence: there was only a slight intensity difference in the medial position, but a stronger effect in the final position. Sluijter and van Heuven (1996) showed that listeners used intensity as a cue to detect word-stress position, but to a lesser degree than, for example, duration. In an important paper, Pierrehumbert (1979) studied intensity (amplitude) in two of the perception experiments. She found that the amplitude effect was 1.5 Hz/dB with regard to the so-called crossover point, where the two f_0 peaks were perceived as equally prominent. That is, the increased amplitude during the last peak increased its prominence so that the crossover point was lower by 1.5 Hz for each increased dB in amplitude. She concluded that while intensity plays an important role in the perception of prominence, its effect does not match the effect of f_0 in importance. How much of this holds for Finnish, is to be determined.

2. Experiments

Experiments 1 and 2 were conducted in order to investigate the perception of prominence in Finnish. The first experiment laid the basis for the tonal features whereas the second one was used to investigated the influence of syntactic structure, namely, word order, on the perception of prominence.

In Experiment 1, the sentence "Menemme laivalla Lemille" (We go by boat to Lemi) 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] by 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?").

Only the first three conditions were investigated in the current study. This was done in order to keep the experimental setup sufficiently simple, but informative enough, for the investigation of both the tonal information underlying the perception of prominence within the experiments as well as the influence of word order on that perception between the experiments.

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 Fig. 1 in Section 2.1.

We employed the Fujisaki model (Fujisaki & 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 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 a set of stimuli to be manipulated effectively 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, Vainio, Werner, & Järvikivi, 2002; Vainio et al., 2003).

The peak heights in Table 1 are in Hertz for the reader's convenience. Since the temporal differences between the peak beginnings and maxima are constant, the three possible values (namely peak magnitude in Hz, semitones, and Fujisaki accent command amplitudes) are in complete correlation. That is, they all implicitly stand for the rate of change. Moreover, since we are dealing with a curve rather than a straight line, an approximation of the rate of change is the most practical measure for the subsequent statistical analyses, which are based on semitones in

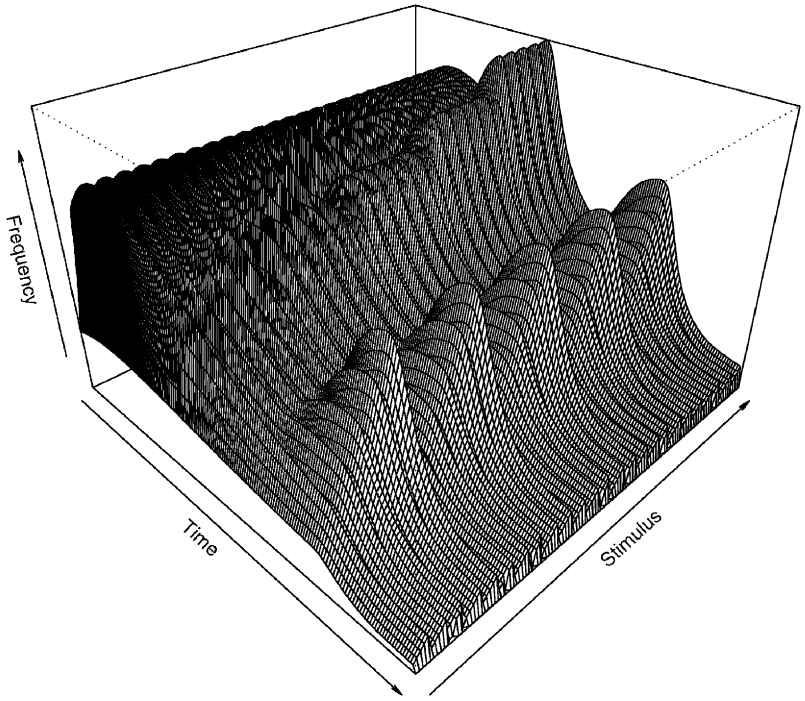

Fig. 1. A three-dimensional view of the 125 f_0 contours of the stimuli used in the experiments. The contours are ordered so that each declination type is cycled over, then the second peak over the declination type and finally the first peak over declination type and second peak. The first stimulus, therefore, has a low declination value and no accents and the last stimulus has high declination as well as high accent values.

Table 1
Declination and peak values in Experiment 1

Declination		Peak 1		Peak 2	
ap	st (per second)	aa1	Mean (Hz)	aa2	Mean (Hz)
1.2	7.7	0.46	32.6	0.46	27.0
1.0	6.4	0.34	20.6	0.34	16.8
0.8	5.0	0.23	9.7	0.23	7.7
0.6	3.6	0.11	0.0	0.11	0.0
0.4	2.9	0.0	−9.0	0.0	−8.0

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 denote the 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.

our case. One point to note about the Table 1 is the fact that when there is no accent command in the Fujisaki model, the peak has a negative value. This is due to the fact the values in the table are calculated from the actual f_0 contours using the time points shown in Fig. 1; the lack of an accent simply shows the underlying declination during the given time that the accent rise would take place. The negative values have no unintended consequences in the analyses.

2.1. Experiment 1

2.1.1. Participants

Twelve phonetically untrained students from a linguistics 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 both the naturalness of the utterances and how well an utterance represented 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, the utterance was selected from a set which was unanimously judged to belong to the intended category of broad focus; no single word was perceived to be more prominent than others. The speaker was a 41-year-old male (the first author) from Helsinki.

The original utterance from Mixdorff et al. (2002) that was used as a basis for stimulus construction in the present study 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 Fig. 1 for more detail).

The Fujisaki model, being superpositional, models the accent peaks as responses to step functions, whose onsets determine the point where the rise begins, whose amplitudes determine the height of the peak, and whose offsets determine where the fall is located. With these commands it is possible to model the accents very accurately with regard to both their temporal structure and their magnitude. In this study, the so-called beta constant of the model was set in such a way that the produced curve always reached the value set by the accent command amplitude. The f_0 curve was modeled as closely as possible with regard to the accent beginnings and ends, and the timing of the accents was kept constant with regard to the accented syllables. The timing was determined by the original stimulus. Therefore, in the resulting stimuli, only the peak magnitude was varied relative to the baseline declination determined by the Fujisaki model phrase component. The accent onsets coincided roughly with the syllable onsets and the peak maxima occurred at the end of the stressed vowel in the word "Lemille" and in the middle of the stressed diphthong in the word "laivalla". This reflected very accurately the original f_0 contour. The timing of the accents was also in line with the results of Suomi, Toivanen, and Ylitalo (2003), who found that in Finnish accentuation follows a moraic pattern, where the rise occurs during the first mora of the syllable and the fall on the second one—a diphthong in Finnish is considered to consist of two morae.

The segmental durations of the stimuli were also kept constant. The durations of the original stimulus reflect those of moderately accented syllables. As Suomi and Ylitalo (2004) showed, all lexically stressed syllables in content words are lengthened, whereas strongly accented syllables are lengthened more than moderately accented or deaccented ones. Therefore, the stimuli used in this study were controlled with regard to durational cues.

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 s and peaked at 0.72 s; the second peak rise started at 0.99 s and peaked at 1.23 s. There was always a fall between the peaks. The utterance ended with a vocal fry during the last two syllables of the utterance.

A set of parameterized 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 Fig. 2. The resulting pitch contours were superimposed on the original utterance with a time-domain PSOLA (pitch-synchronous overlap and add) method. As mentioned above, 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.[1] Each block was preceded by a second long sine tone of around 400 Hz, and a 15 s pause was inserted between each

[1]The baseline stimulus with the parameter values of the original utterance was included in each stimulus block in order to check any possible effects caused by hearing the same utterance multiple times.

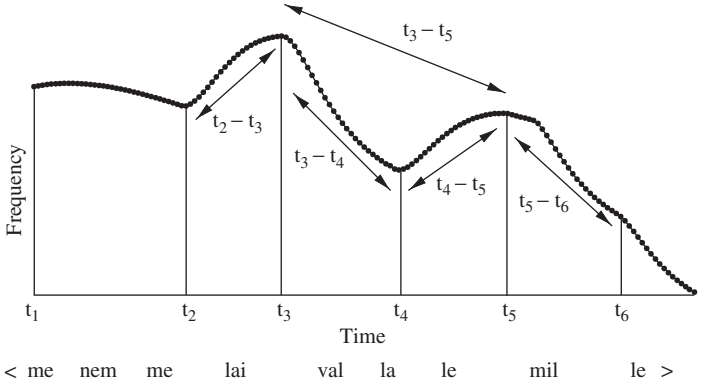

Fig. 2. An example f_0 contour with an orthographic transcription separated into syllables. The points of interest in the contour are marked with t_n. The double-ended arrows depict the actual factors used for statistical analyses. Note that all the analyzed values are in f_0 corresponding to the changes in the pitch countour at given times.

block. The inter-stimulus interval was set to 4 s. Ten practice trials preceded the first experimental trial. The stimuli for the practice trials were randomly chosen from the experimental materials. The purpose of the practice trials was to familiarize the subjects with the procedure and to make sure that they understood the instructions.

A ternary decision instead of a simple two alternative forced choice procedure (2AFC) was used. 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.[2] 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 judgments based on relative prominence which is the result of various factors besides pitch alone.

Fig. 2 shows the basic f_0 contour used in the experiments and lists the points of interest (from t_1 to t_6) and the dynamic factors (attained by subtracting the frequency values in semitones between any given two points; $t_i - t_j$) used in subsequent statistical analyses. For instance, the factor $t_2 - t_3$ stands for the amount of rise of the first peak.

2.1.4. Results and discussion

The effect of each operationalized factor was determined by regression analysis. We further assessed the effects of the baseline declination and of the peak sizes on the subjects' responses with analyses of variance (ANOVAs).

[2]Since the stimuli naturally fall into three categories, we also wanted to keep the choices as natural as possible, thus avoiding having to force the subjects into making a binary decision when the input clearly does not support such an assignment. Moreover, selecting "on neither" as the response category 3 was motivated by a pilot study where the third category was instead "on both": the results showed a large number of random assignments that were interpreted to be due to individual strategies when the acoustic cues were the least salient, i.e., when they did not support any of the three choices.

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 in f_0 between the two peak heights; $t_3 - t_5$ in Fig. 2 ($t = 12.30$, $p < 0.001$) and the amount of rise in the first peak; $t_2 - t_3$ in Fig. 2 ($t = -5.24$, $p < 0.001$). The overall regression model with the above two factors was highly significant; [$F(2, 127) = 182.0, R^2 = 0.74, p < 0.001$]. 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 ($t = -11.05$, $p < 0.001$) and the fall of the last peak, i.e., the difference in f_0 between points t_5 and t_6 ($t = 6.51$, $p < 0.001$). Again, the overall regression model was highly significant [$F(2, 127) = 191.9, R^2 = 0.75, p < 0.001$]. 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 important feature modulating the perception of prominence for both Responses 1 and 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 from what could also be expected intuitively, that is, the magnitude of rise.

Moreover, the stress was perceived unanimously to be on the last word, when the first peak was lower than the second one in absolute terms. We will return to this in Section 2.2.4 as well as in the general discussion.

To assess the effects of the manipulation of the phrase component (leading to changes in baseline declination, hence Phrase) and of the 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 Responses 1 and 2 categories with Accent standing for the first and second pitch peak, respectively.

Response 1: ANOVAs revealed a significant main effect of Phrase [$F(4, 44) = 8.61, p < 0.001$] as well as Accent [$F(4, 44) = 107.24, p < 0.001$]. There was also a significant interaction of Phrase and Accent [$F(16, 176) = 1.92, p < 0.05$].

Response 2: There was again a significant main effect of Phrase [$F(4, 44) = 9.26, p < 0.001$] as well as Accent [$F(4, 44) = 76.03, p < 0.001$]. There was also a significant interaction of Phrase and Accent [$F(16, 176) = 2.71, p < 0.01$].

The main effects of Accent in both response categories are obvious in light of the regression analyses 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 indicates 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 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 since 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 Fig. 3). However, the perception of prominence may be attributable to other than purely phonetic factors, such as focus and word order. Thus, Experiment 2 investigates whether placing the focus on the last word by changing the order of the two nouns in the adverbial phrase affects the prominence judgments as compared to Experiment 1. In other words, whether non-phonetic factors, here word order, have consequences as to how prominent the last peak is perceived to be.

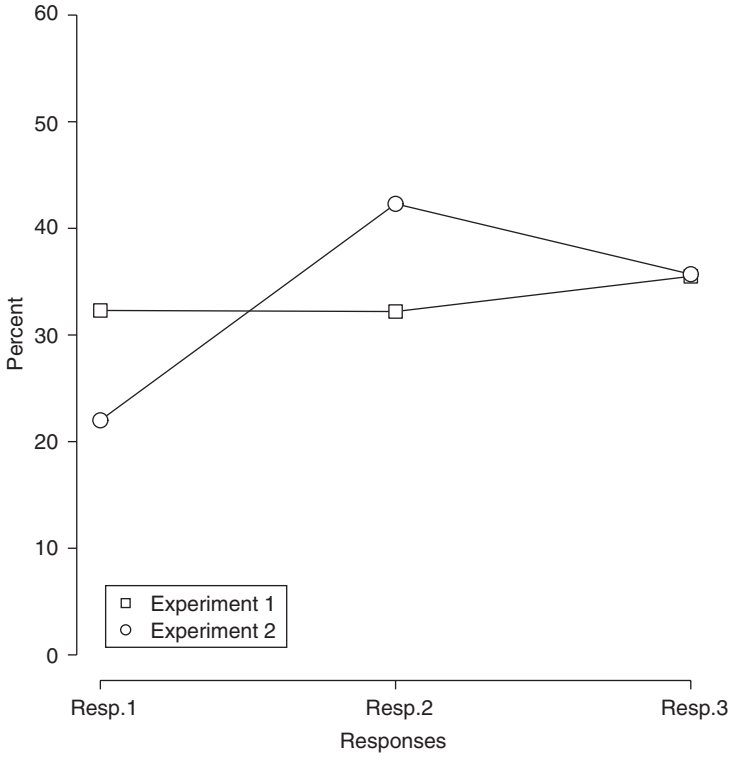

Fig. 3. Distribution of responses in Experiments 1 and 2. The squares depict Experiment 1 and the circles Experiment 2.

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 & 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, ''by what means are you going to Lemi?'', the phrase ''menemme Lemille laivalla'' would translate into English as ''it is by boat we are going to Lemi''. 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 the word order manipulation also to affect the subjects' prominence judgments. 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 research and none had participated in the first experiment.

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 a 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 high-quality analog-to-digital transformer. The recording was done with 16 bit quantization using 44.1 kHz sampling frequency. The utterance was then subjected to the same procedure as described in the previous experiment, and a corresponding 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 2.2.4 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

Fig. 3 shows the distribution of responses with regard to the sentence stress conditions 1–3 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 (32.3%, 32.2%, and 35.5%, respectively), there were more responses for the second condition (stress on the last word) in Experiment 2 (42.3%) and fewer responses for the first condition (22.0%), whereas, notably, the proportion of responses in category 3 (35.7%) did not

differ from that in Experiment 1. A χ^2-test on the distributions showed a significant difference between the experiments ($\chi^2(2) = 47.34$). The change in the distribution of judgments observed in Experiment 2 suggests that the participants' perception of prominence was indeed affected not just by the purely tonal factors, but also by higher-order linguistic information. In other words, the results show that the manipulation of the syntactic (and information) structure causing the focus to be placed on the last word, also attracted more prominence judgments on the last peak, despite the fact that the subjects were explicitly instructed to make decisions about the stress in each sentence. More importantly, the placing of focus on the last peak affected the prominence judgments of the first peak to a similar degree, although in an opposite direction.

We further analyzed the responses with regard to the top-line declination to see how the word order change had influenced the relative prominence of the two adverbials. Fig. 4 shows the distribution of both responses 1 and 2 in the first two experiments vs. the absolute peak difference of the stimuli. As there were three variables influencing the responses, a simple line showing the distribution cannot be drawn. Instead we have used locally estimated regression lines to show the differences between the experiments. These lines are not interpretable statistically and we, therefore, estimated the crossover points with probit analysis (Venables & Ripley, 1996) for both responses in the experiments by pooling the responses for all participants. The values for Response 1 are −31.6 Hz (Experiment 1), and −40.6 Hz (Experiment 2) indicating that the first peak has to be raised an additional 10 Hz to be perceived with the same prominence as the second peak when the word order is changed. Similarly, the results for Response 2 (Experiment 1, −5.7 Hz, and Experiment 2, −13.7 Hz) indicate that the last peak was perceived to be as prominent as the first when the declination was an additional 8 Hz steeper. That is, with both responses there was a clear bias which either resisted the responses from being categorized as the first peak being more prominent, or attracted more responses to the last peak when the word order was changed from unmarked to marked (see Fig. 3). Furthermore, the bias was fairly evenly distributed throughout the accent space provided by the stimuli. The crossover points are further discussed in Section 3.

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 difference between the two peak heights $t_3 - t_5$ ($t = 11.33$, $p < 0.001$); and the amount of rise in the first peak, $t_2 - t_3$ ($t = 3.76$, $p < 0.001$). The overall regression model was highly significant [$F(2, 127) = 131.6$ and $R^2 = 0.67$, $p < 0.001$]. 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 ($t = -10.79$, $p < 0.001$) and the fall of the last peak ($t = 8.37$, $p = 0.001$). Again, the overall regression model was highly significant [$F(2, 127) = 241.1$ and $R^2 = 0.79$, $p < 0.001$]. 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 had a similar structure from a perceptual point of view: the most important feature for the first peak was its rise while the last peak was characterized by a fall. With regard to the actual contrast between the peaks, the difference between the peak heights was, again, by far the most important factor. Moreover, the sentence stress was almost

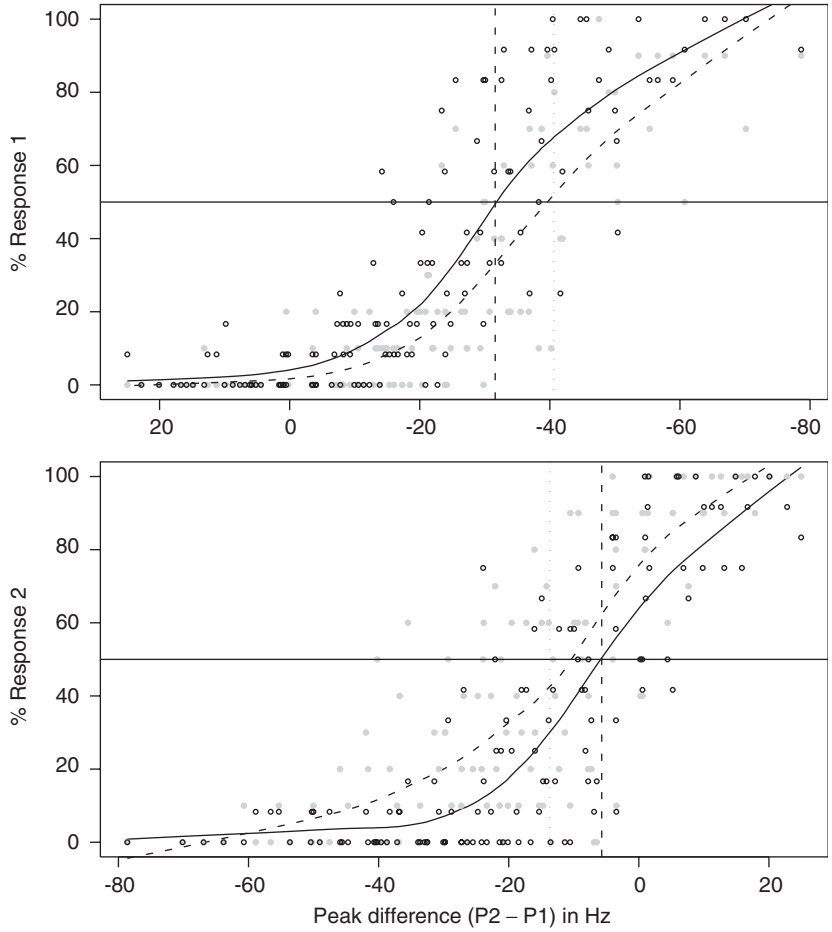

Fig. 4. Distribution of responses (y-axis) in % vs. the absolute difference between the two peaks (x-axis) for Experiments 1 and 2. The upper panel shows the results for Response 1. The gray dots in both panels depict the individual values for each stimuli in Experiment 1 whereas the circles stand for results in Experiment 2. Locally estimated regression lines are also drawn for each experiment: solid line for Experiment 1, dotted line for Experiment 2. The lower panel depicts results for Response 2. The crossover points provided by the probit analyses are marked with dotted (Experiment 2) and dashed lines (Experiment 1). Note that for making the plots easier to interpret, we have reversed the x-axis of the upper plot.

unanimously perceived to be on the last word when the first peak was lower than the second one in absolute terms. However, the phenomenon was not as pronounced as in Experiment 1.

With respect to the effects of the phrase and accent components, the results are as follows: for Response 1, ANOVAs revealed a significant main effect of Phrase [$F(4, 36) = 10.08, p < 0.001$] as well as of Accent [$F(4, 36) = 47.37, p < 0.001$]. There was also a significant interaction between Phrase and Accent [$F(16, 144) = 1.99, p < 0.05$]. As to Response 2, there was again a significant main effect of Accent [$F(4, 36) = 137.99, p < 0.001$]. However, there was no main effect of Phrase [$F(4, 36) = 2.20, p > 0.08$]. There was also a significant interaction of Phrase and Accent [$F(16, 144) = 1.98, p < 0.05$].

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, and hence 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.

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 therefore not designed to be symmetrical in the sense that both potentially accented syllables would have the same structure. The stimulus 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 sentence.

The second word-order condition, however, is not neutral with regard to focus. The focusing function provided by the marked word order could influence the relative prominence of the constituents. That is, the speaker might compensate negatively for the prosodic prominence due to the added salience provided by the marked word order. Therefore, we could not use a similar procedure as was used for the first experiment to select a candidate utterance. The basic stimulus for the experiment was chosen from a set of utterances that the speaker clearly intended to produce with prosodically as neutral or broad a focus as possible. The post hoc analysis, however, showed that the manipulation of word order affected some of the phonetic characteristics of the utterance. First, the relative intensity between the two accent peaks was different in Experiment 2 than 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 discount the slight possibility that the observed difference in Experiments 1 and 2 was affected by these factors, we designed two further experiments to investigate the role of the degree of fall in the last accent peak (Experiment 3) and the role of intensity (Experiment 4).

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 intensity 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 (due to the different segmental make-up of the accented syllables in the words "laivalla" and "Lemille") 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 the syllable's 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 pattern with those of the second experiment. If, however, the small difference in the degree of the fall had no effect, the results should be consistent with those of the first experiment.

As noted above, the relative intensity between the two accented words in the two experiments was different in that the last noun (''laivalla'') 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 Fig. 5.

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 to the difference in word order. If that was indeed the case, the perceptual advantage gained from the added intensity in the second experiment should be weakened 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 1 and 2.

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 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

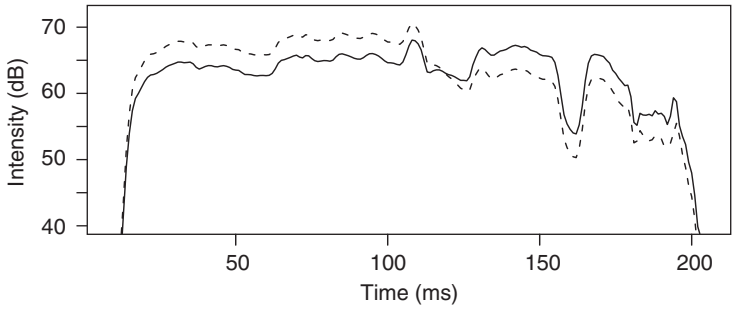

Fig. 5. Intensity curves of the baseline stimuli in Experiment 2 (solid line) and Experiment 4 (dotted line).

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

Considering firstly 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 χ^2-test on the proportions showed a non-significant difference between the results ($\chi^2(2) = 4.85$, $p = 0.78$). 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.

Turning now to the effect of intensity, Fig. 6 shows the distribution of responses in Experiment 4 against the responses in Experiments 1 and 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 last word. What is remarkable is that, while the responses for the last 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 last peak and has no global effect.

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

As the only difference between Experiments 2 and 4 was the relative intensity between the last of the two accented nouns, we analyzed the effect of intensity by pooling the responses of both experiments for both linear regression analyses and repeated measures ANOVAs. When added as a regressor, intensity was significant with respect to Response 2 ($t = 4.33$, $p < 0.001$), 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 experiments 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 [$F(1, 22) = 6.20, p < 0.05$] as well as a significant interaction between Experiment and Accent in Response 2 [$F(4, 88) = 2.94, p < 0.05$]. This clearly shows, that

Table 2
The distribution of responses in all four experiments as percentages

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Response 1	32.3	22.0	33.4	24.3
Response 2	32.2	42.3	32.5	33.6
Response 3	35.5	35.7	34.1	42.1

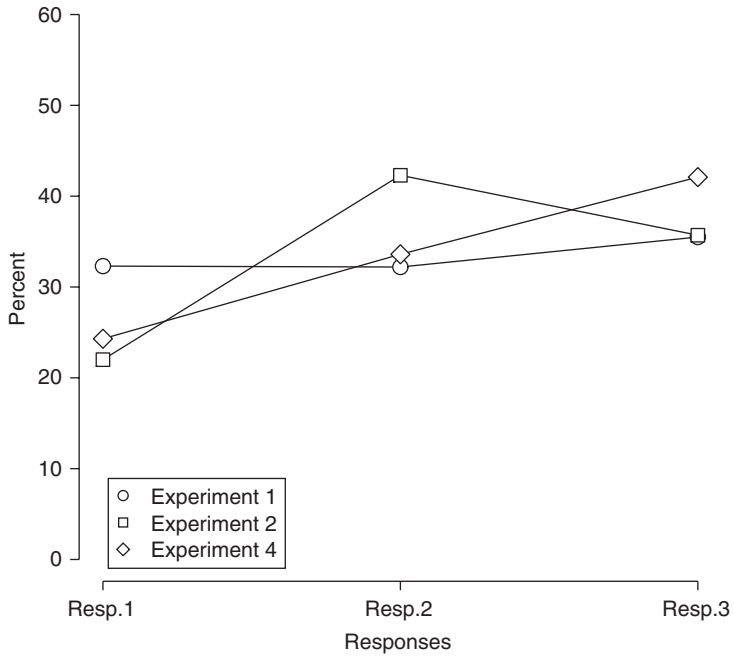

Fig. 6. Distribution of responses in Experiments 1, 2, and 4. The squares depict Experiment 1, circles Experiment 2, and diamonds Experiment 4.

the effect of intensity is local to the last 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 is mostly due to increased intensity of the last 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. That is, there appears to be a perceptual trade-off between intensity and pitch height which the speaker can take advantage of.

3. General discussion

The present article has reported the results from 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. Our aim, then, was to investigate 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 last 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 last 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 a phonetic phenomenon as stress.

At this point the possible consequences of adopting a three-choice procedure rather than the classic two-alternative forced choice method should be briefly discussed. Schouten, Gerrits, and van Hessen (2003) showed that the two-choice paradigm introduces a strong subjective response bias in classic phoneme discrimination (categorical perception) experiments. Moreover, they showed that the effect of individual response strategies was largest on either side of the phoneme boundary. Thus, the less information there is in the input helping to cue the choice, the more the choice depends on individual strategies. More importantly, however, they also showed that the subjective bias is considerably lessened when the task does not force the participants to make a choice in the absence of sufficient cues to guide the response. In the present experiment, the adoption of a three-choice method, rather than 2AFC, can be argued to have been less open to subjective bias, and thus to random effects, than the two-choice method. The experimental stimuli also fell naturally to the three focus categories corresponding to the choices of response, thus considerably reducing the need to make a choice that was clearly not supported by the input. The evidence from Experiments 1, 2, and 4 clearly supports this conclusion: while the word-order change in Experiments 1 and 2 affected only the distribution of the Responses 1 and 2 (as they should), there was no effect between Experiment 2 and 4 on the Response 1 category. The intensity manipulation was instead realized in the difference in distributions between categories 2 and 4, suggesting indeed that the effect of intensity was strongest in sentence final position, a result which is well in accordance with what Heldner (1996) found to be the case in production. However, it is possible that the effect of word order could have come out more strongly with the adoption of 2AFC, since it could be argued that when the tonal information was not salient enough to provide sufficient cues for the decision, the participants could have relied more on the word order.[3] However, it is also true that since tonality was clearly stronger in guiding the responses than the word order, a lack of acoustic support would most likely have increased the uncertainty and thus the need to base the response on subjective criteria. Rather than highlighting the effect of word order, the potential random effect caused by the increase of response bias could have in fact wiped out the entire effect. Now this danger was avoided by giving the subjects a natural way out by choosing category 3. Since adopting an admittedly more conservative method was nevertheless enough to bring out the effect clearly, nothing could have been gained by using the 2AFC instead.

[3]This possibility was pointed out to us by one of the anonymous reviewers.

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

- The first f_0 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 perceived prominence of the first f_0 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 the 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 findings from other languages. For instance, the fact that the last 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 for 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, Werner, Vroomen, & de Gelder, 1999; Tuomainen, 2001). The "sagging transitions" are also in line with similar findings for English by Ladd and Schepman (2003).

As mentioned before, the role of relative intensity in the perception of prosody is a fairly uncharted territory. Nevertheless, Pierrehumbert (1979) did study intensity (or amplitude) in a comparable setting to ours. As mentioned in Section 1.2, 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 last 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 last 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.

Table 3
The crossover points for responses 1 and 3 in all experiments in Hertz and semitones followed by Standard Errors in parentheses

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Response 1 (Hz)	−31.6 (0.83)	−40.6 (1.10)	−31.1 (1.02)	−39.5 (0.98)
Response 1 (Semitone)	−5.51 (0.19)	−6.93 (0.16)	−5.41 (0.17)	−6.77 (0.15)
Response 2 (Hz)	−5.7 (0.79)	−13.7 (0.84)	−5.2 (1.03)	−8.3 (0.67)
Response 2 (Semitone)	−1.06 (0.13)	−2.54 (0.14)	−1.03 (0.17)	−1.58 (0.11)

Declination is, in fact, considered to be a universal phenomenon and there is even evidence that it occurs in certain monkey species' vocalizations (Hauser & Fowler, 1992). It is therefore interesting to see how the phonetic and linguistic factors studied here influenced the crossover points in the experiments.

The estimated crossover point values for all experiments are summarized in Table 3. Whereas the values for Response 2 are in line with Pierrehumbert's results (see below), the values for Response 1 are much lower.[4] This can be explained by the presence of baseline declination in the stimuli (Pierrehumbert varied the magnitude of the last peak only). The top-line difference in itself is not a sufficient cue for prominence; a local rise for the accent must be added to the baseline in order for the word to be perceived as stressed.

According to Pierrehumbert (1979), the difference between the points in Experiments 1 and 2 should have been approximately 8 Hz (for an approximately 5 dB increase in intensity), which is exactly the value we obtained for Response 2 (the crossover points for Experiments 1 and 2 being −5.7 Hz, SE = 0.7934 and −13.7 Hz, SE = 0.8432, respectively). However, when we look at the results for Experiment 3, we do not get such a marked difference (−8.3 Hz, SE = 0.6728). These results 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 (−5.2 Hz, SE = 1.0279 as opposed to −5.7 Hz). In summary, there is a clear intensity effect with regard to Response 2, whereas for Response 1 none can be observed. In contrast, the differences between the crossover points for Response 1 show almost identical values for Experiments 1 and 3 as well as for Experiments 2 and 4. Moreover, the values for Experiments 2 and 4 are almost 10 Hz higher than the values for 1 and 3. The differences are clearly attributable to one factor only, namely word order. It therefore seems, that in the cases with marked word order, as in Experiments 2 and 4, the peak had to be an additional 10 Hz higher than in the unmarked cases in order for the first word to be perceived as more prominent than the second.

[4]The fact that the crossover points are not symmetrical with respect to the word position further increases the methodological value of using a three choice paradigm. A 2AFC paradigm would have forced a non-existent symmetry and, thus, skewed the results towards a non-existent mean.

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 Experiments 1 and 2 as well as on 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 cognitive 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 also to abstract linguistic characteristics of the utterance.

That Finnish listeners should be sensitive to both acoustic and syntactic cues in the perception of prominence, is perhaps not surprising considering that in Finnish syntactically free word order nevertheless serves systematic pragmatic information structure functions such as focusing. Since intonation and syntactic structure are the most important vehicles of information structure, the word order affects the perceived prominence via its role in focus placement in Finnish. Would this be the case in other, syntactically different and similar, languages? Donati and Nespor (2003) argue that intonational focus and syntactic structure tend to be related: the more prominence can move around in the intonational phrase the more rigid the word-order properties of the language are. Thus, on the one hand, the role of prosodic prominence for focus placement is very high in languages where word order cannot be used for such functions. For example, Swerts, Krahmer, and Avesani (2002) showed that information status is clearly reflected in perceived prominence differences in Dutch, whereas there was no straightforward connection between the two in Italian, a language with a freer word order. Thus, it could be hypothesized that syntactic structure would have less effect on the perception of prominence in languages with more rigid word order than in Finnish. On the other hand, there is evidence that free word-order languages with pragmatically less strict constraints may also have consequences with respect to sentence processing. Bojar, Semecky, Vasishth, and Kruijff-Korbayova (2004) showed in a reading study that in Czech, a language with pragmatically less constrained and higher-frequency non-canonical word-orders, the processing cost induced by word-order changes was not as high as in Finnish, where under similar circumstances non-canonical word order, namely OVS, has been shown to induce processing difficulties compared to the canonical SVO word order in the absence of a facilitating context (Hyönä & Hujanen, 1997; however, cf. Kaiser & Trueswell, 2004). Although, the available evidence is only indirectly related to the issue at hand, it might be hypothesized that the relative influence of linguistic structure on the subjective prominence assignment is at its greatest with languages, such as Finnish, where the trade-off between intonation and syntactic structure for focus placement is sufficiently large due to well-defined pragmatic functions of the word-order changes. By contrast, its influence might be lessened both in languages with either syntactically more constrained word order or syntactically free but pragmatically less-constrained word order. At the moment, however, this is hardly more than speculation, and the issue must be left for future research.

In sum, 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 and duration of the syllables and the syntactic structure represented by word order.

Acknowledgements

We would like to thank Rachael-Anne Knight and three anonymous reviewers as well as Jukka Hyönä, Pirita Pyykkönen and Hanna Westerlund for their insightful comments on the manuscript. We also thank Stefan Werner and Hansjörg Mixdorff for their contributions to the discussions on this subject. 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

Batliner, A., Buckow, J., Huber, R., Warnke, V., Nöth, E., & Niemann, H. (2001). Boiling down prosody for the classification of boundaries and accents in German and English. In *Proceedings of the European conference on speech communication and technology*, September 2001, Aalborg (Vol. 4, pp. 2781–2784).

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustical Society of America*, *67*(2), 648–662.

Bojar, O., Semecky, J., Vasishth, S., & Kruijff-Korbayova, I. (2004). Processing noncanonical word order in Czech, a poster presented at *Architectures and Mechanisms for Language Processing*, September 2004, Aix-en-Provence, France.

Donati, C., & Nespor, M. (2003). From focus to syntax. *Lingua*, *113*, 1119–1142.

Eriksson, A., Thunberg, G. C., & Traunmüller, H. (2001). Syllable prominence: A matter of vocal effort, phonetic distinctness and top-down processing. In *Proceedings of the European conference on speech communication and technology*, September 2001, Aalborg (Vol. 1, pp. 399–402).

Fujisaki, H., & Hirose, K. (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan (E)*, *5*(4), 233–241.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, *102*(5), 3009–3022.

Gussenhoven, C., & Rietveld, T. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics*, *16*, 355–369.

Hakulinen, A., & Karlsson, F. (1979). *Nykysuomen lauseoppia*. Helsinki: Suomalaisen Kirjallisuuden Seura [Finnish Syntax].

Hauser, M., & Fowler, C. (1992). Fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates. *Journal of the Acoustical Society of America*, *91*, 363–369.

Heldner, M., 1996. Is an F0 rise a necessary or a sufficient cue to perceived focus in Swedish? In S. Werner (Ed.), *Nordic Prosody: Proceedings of the VIIth conference*, Peter Lang (pp. 109–125).

Hermes, D. J. (1997). Timing of pitch movements and accentuation of syllables in Dutch. *Journal of the Acoustical Society of America*, *102*(4), 2390–2402.

Hume, E., & Johnson, K. (2003). The impact of partial phonological contrast on speech perception. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 2385–2388). Spain: Barcelona.

Hyönä, J., & Hujanen, H. (1997). Effects of word order and case marking on sentence processing in Finnish: An eye fixation analysis. *Quarterly Journal of Experimental Psychology*, *50A*, 841–858.

Kaiser, E., & Trueswell, J. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

Ladd, D., & Schepman, A. (2003). "Sagging transtitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, *31*, 81–112.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, *25*, 313–342.

Ladd, D. R., Verhoeven, J., & Jacobs, K. (1994). Influence of adjacent pitch accents on each other's perceived prominence: Two contradictory effects. *Journal of Phonetics*, *22*, 87–99.

Mixdorff, H., Vainio, M., Werner, S., & Järvikivi, J. (2002). The manifestation of linguistic information in prosodic features of finnish. In B. Bel, I. Marlien (Eds.), *Proceedings of Prosody 2002* (pp. 515–518).

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, *66*, 363–369.

Remijsen, B., & van Heuven, V. J. (2003). On the categorical nature of intonational contrasts. In J. van de Weijer, V. J. van Heuven, & H. van der Hults (Eds.), *The phonological spectrum. Amsterdam studies in the theory and history of linguistic science*, John Benjamins Publishing Company (Vol. II, pp. 225–246).

Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474–494.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, *100*(4), 2471–2485.

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

Suomi, K., & Ylitalo, R. (2004). On durational correlates of word stress in Finnish. *Journal of Phonetics*, *32*, 35–63.

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, *30*(4), 629–654.

Terken, J. (1989). Reaction to C. Gussenhoven and A. Rietveld: Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics*, *17*, 357–364.

Terken, J. (1994). Fundamental frequency and perceived prominence of accented syllables, II: Non-final syllables. *Journal of the Acoustical Society of America*, *95*, 3662–3665.

Terken, J., & Hermes, D. (2000). The perception of prosodic prominence. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 89–127). Dordrecht: Kluwer Academic Publishers.

Tuomainen, J., 2001. Language specific cues to segmentation of spoken words in Finnish: Behavioral and event-related brain potential studies. Ph.D. thesis, de Katholieke Universiteit Brabant.

Tuomainen, J., Werner, S., Vroomen, J., & de Gelder, B. (1999). Fundamental frequency is an important acoustic cue to word boundaries in spoken Finnish. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A.C. Bailey (Eds.), *Proceedings of the 14th congress of phonetic sciences*, San Francisco (pp. 921–924).

Vainio, M., Mixdorff, H., & Järvikivi, J. (2003). Perception and production of focus in Finnish. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences*, Barcelona, Spain (pp. 1831–1834).

Van Valin, R., & La Polla, R. (1997). *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.

Venables, W., & Ripley, B. (1996). *Modern applied Statistics with S-plus*. Berlin: Springer.

Vilkuna, M. (1989). *Free Word Order in Finnish: Its Syntax and Discourse Functions*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research* 7, 357–369.