The Production and Perception of Focus in Finnish

Martti Vainio∗, **Hansjörg Mixdorff**∗∗ **Juhani Järvikivi**† **Stefan Werner**‡

∗ University of Helsinki, Department of Phonetics and Department of General Linguistics
`martti.vainio@helsinki.fi`

∗∗ Berlin University of Applied Sciences, Germany
`mixdorff@tfh-berlin.de`

† University of Dundee
`juhani.jarvikivi@utu.fi`

‡ University of Joensuu
`stefan.werner@joensuu.fi`

Abstract

This paper presents a study examining the production and perception of focus in Finnish. In a corpus of segmentally identical utterances the F_0 contour was systematically varied to elicit different perceptions of focus. All utterances were rated with respect to their perceived naturalness by 12 native speakers of Finnish. Based on these results an experiment with respect to the perception of focus using re-synthesized stimuli was performed. In the test the subjects had to decide which of the four possible focus conditions they perceived. The results include, inter alia, that the second accent in the utterance must be raised by at least 2.6 semitones from the baseline to be perceived to have narrow focus on the latter word of two-accent utterance. Stimuli with rises closest to the means given for the four conditions were generally identified as belonging to the intended category, though the condition "broad", apparently the default choice, covers a large triangular region in the two-dimensional accent space. For the broad condition to be unanimously perceived also requires that the latter accent peak be lower than the first – this gives indirect evidence to the hypothesis that Finnish listeners and speakers normalize for declination.

1 Introduction

The linguistic phenomenon of focus can be achieved by various syntactic and prosodic means, for example, by increasing the prominence of the part of an utterance that is intended to be brought into focus. Thus, focus as a linguistic means is closely related to (the perception of) prominence. Finnish can allow any word in an utterance to be focused by prosodic means: thus a Finnish speaker can say "*Manne* meni Lemille"

(English: "*Manne* went to Lemi") as well as "Manne meni *Lemille*" (English: "Manne went to *Lemi*" (italics depict focus). The free word order also allows for focusing by simply placing the focused word at the end of the utterance, thus "menemme Lemille laivalla" as opposed to "Menemme laivalla Lemille" ("We go to Lemi by boat" vs. "We go by boat to Lemi") with an unmarked word order can be expected to have the focus on the last word "laivalla" – as long as the word "Lemille" is not accented.

This serves as a good opportunity to study the relation between the accent strength and the excursion size of the accents in a multiple accent utterance and their relation to focus. This study falls in line with a series of somewhat similar studies reported by e.g. Pierrehumbert [1], Gussenhoven and Rietveld [2], Terken [3] and, Ladd et al [4], as well as Gussenhoven et al [5]. The main difference is that the earlier studies concentrated mainly on prominence, whereas here we are interested in focus with indirect implications to prominence.

The current paper presents a study of Finnish prosody using the Fujisaki model [6] and concentrates solely on the tonal aspects of the language. It is a continuation of our earlier work on the suitability of the Fujisaki model to parametrize Finnish intonation [7]. Although the Fujisaki model plays a pivotal role in this study, we did not attempt to evaluate the model per se, but utilized it as a means to produce phonetically constrained stimuli for perception experiments. Therefore, it makes more sense to present the results on a perceptually relevant scale rather than parameter values for the model. For this we have chosen the semitone scale. The merits of using the semitone scale are argued for and against in the literature (see, for instance [8] and the references therein). Nevertheless, our results do not require such accuracy that would merit the use of any other scale, such as the ERB scale.
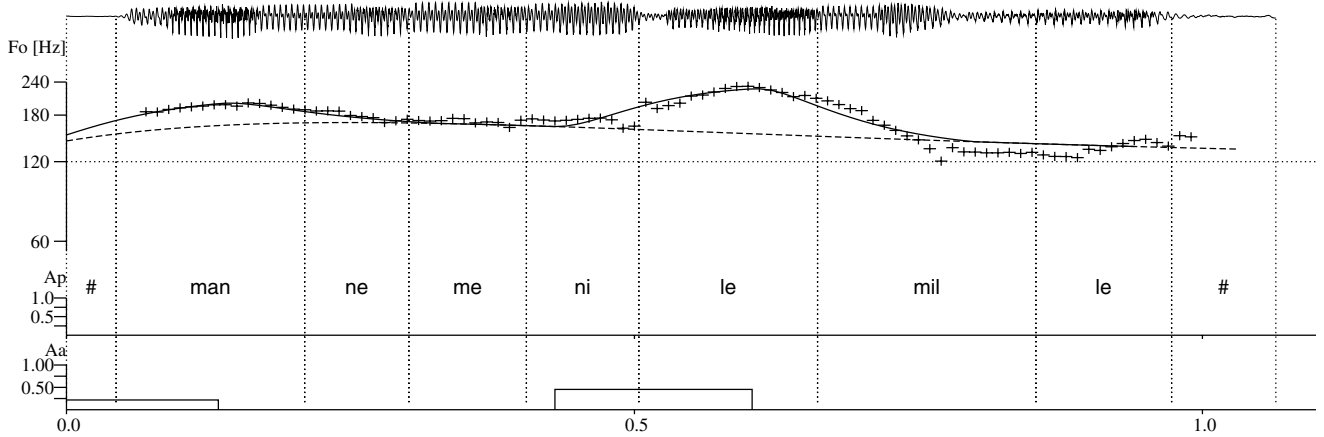
Figure 1: A typical Finnish utterance "Manne meni Lemille" with a narrow focus on the word "Lemille" produced by a female speaker of Finnish and analyzed with the Fujisaki model. See text for more detail.

We used the sentence "Manne meni Lemille" (Manne went to Lemi) which permits four possible interpretations with respect to focus:

1. broad

2. narrow focus on "Manne"

3. narrow focus on "Lemille"

4. multiple contrastive, narrow focus on both "Manne" and "Lemille"

The modeled F_0 contours for each interpretation are depicted in Figure 2.

A typical utterance spoken by a female speaker and analyzed with the Fujisaki model can be seen in Figure 1. The figure shows the actual waveform of the utterance, the modeled as well as the actual F_0 contours (solid line and +-marks, respectively). The figure also shows a syllabic labeling of the utterance and the parameters for the Fujisaki model (the phrase command is actually outside of the scope of the figure, but the accent commands are conspicuous. The utterance has a narrow focus on the word "Lemille").

2 Speech Material and Method of Analysis

Twelve students participated in the test. The participants were students from different Finnish universities taking part in a workshop on prosody modeling for text-to-speech synthesis. They had a varying background of phonetics and linguistics, but none were specifically trained to transcribe prosody. None of the participants reported any hearing loss. The twelve participants were divided in six pairs, where one acted as

the questioner who's task was to read aloud a prompt sentence to the other participant who then produced the intended reply in the desired focus condition. Table 2 depicts the typical prompt-reply pairs used in the study. The prompt-reply pairs were presented to the participants in written form on a sheet of paper. The dialogues were recorded three times directly on a computer hard-disk using a professional level analogue-to-digital converter and noise-canceling microphones. The recordings were done in a fairly noise free environment at the Department of General Linguistics of the University of Helsinki.

The following table shows the four types of prompt-reply pairs of which the first three consisted in question-answer pairs and the fourth where the prompt was a contrafactual declarative sentence. The Finnish sentences are followed by English translations. Samples of the corresponding F_0 curves can be seen in Figure 2.

A:	Mitä sitten tapahtui?	What happened then?
B:	Manne meni Lemille.	Manne went to Lemi.
A:	Kuka meni Lemille?	Who went to Lemi?
B:	*Manne* meni Lemille.	*Manne* went to Lemi.
A:	Minne Manne meni?	Where did Manne go?
B:	Manne meni *Lemille.*	Manne went to *Lemi.*
A:	Kuulin, että Manu	meni Lemulle.
	I heard that Manu	went to Lemu
B:	Ei – *Manne* meni	*Lemille.*
	No – *Manne* went	to *Lemi*

Table 1: Typical prompt-answer pairs used to elicitate the different focus conditions.

The target utterances were segmented on syllabic level and the parameters for the Fujisaki model were es-
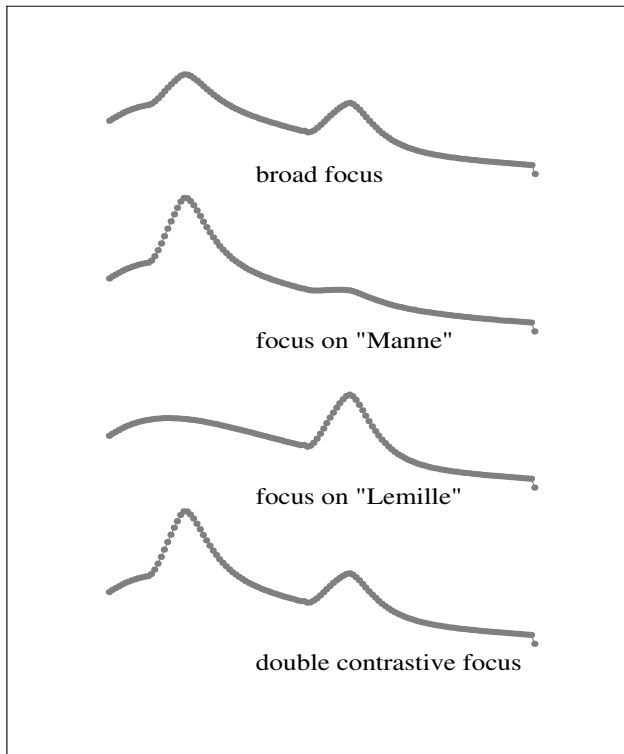
Figure 2: The F_0 contours for the four different focus conditions in the perception test.

timated manually using an interactive program with a graphical user interface. Materials from a typical speaker – in a statistical sense – were selected for further analysis to serve as the basis for the perception experiment.

There was some doubt whether the informants would find it natural and easy to produce the utterances from a written prompt. It turned out, however, that the informants found it very easy to respond to the prompts with an intended focal "meaning". Even the multiple contrast, narrow focus proved to be relatively straightforward to produce when the right context was given. Nevertheless, the double-focus required an extra negation in the form of "ei" ("No"), followed by a pause at the beginning of the utterance.

3 Perception Experiment

The perception experiment was designed to study the influence of accentuation on the intended focal condition.

3.1 Materials
The data from the utterances were analyzed in order to build statistically representative stimuli for the experiment. Typical configurations exhibited one or two accent commands associated with "Manne" and/or

"Lemille". Accent command amplitudes for the four focus conditions are given here for a female speaker whose utterance were later on used in a perception experiment: 1) Aa1/Aa2: 0.25/0.33; 2) 0.41/0.08; 3) 0.00/0.44; 4) 0.42/0.28. The corresponding semitone and Hertz data for the accent rises can be found in Table 2. Based on these results an experiment with respect to the perception of focus using resynthesized stimuli was performed. Starting from an utterance from a broad focus condition, stimuli were created with Aa1 and Aa2 ranging between .00 and .42 in steps of .07 yielding altogether 49 stimuli. The modification was only done with respect to accent command amplitudes; their timing as well as the phrase component were kept constant.

Figure 3 shows the placement of the averaged accent component values for the four different focus conditions mapped onto the accent command amplitude space as stars. The corresponding Hertz and semitone values of the accents can be found in Table 2.

| aa1	p1 st	p1 Hz	aa2	p2 st	p2 Hz
0.00	0.04	0.38	0.00	-0.9	-7.85
0.07	0.78	8.15	0.07	0.03	0.29
0.14	1.52	16.29	0.14	0.84	7.66
0.21	2.26	24.8	0.21	1.71	16.02
0.28	3.0	33.7	0.28	2.58	24.82
0.35	3.74	43.01	0.35	3.45	34.08
0.42	4.47	52.75	0.42	4.32	43.82

Table 2: Accent component values and the corresponding semitone and Hertz values. Aa1 stands for accent component 1 and aa2 for accent component 2, respectively. p1 and p1 stand for the peak heights of the two accents relative to the baseline (phrase component). The values are given in semitones (st) and Hertz (Hz). The Hertz and semitone values were calculated from the actual F_0 contours – the small negative values corresponding to the accent command of 0.0 in p2 are due to the falling phrase component or declination.

3.2 Procedure and Participants
The same twelve students that had participated in the production of the utterances participated also in the perception experiment. The participants were instructed to decide which condition they thought they heard in a forced choice test by marking their response on paper. The list of stimuli were played twice in a random order through good quality loudspeakers in a relatively noisy classroom. The inter-stimulus interval was approximately 2 seconds.

3.3 Results and Discussion
The results include, inter alia, that the second accent in the utterance must be raised by, at least 2.6 semitones from the baseline to be perceived to have narrow focus
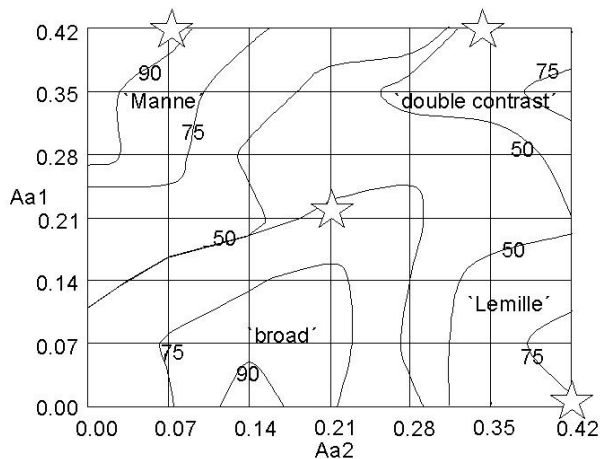
Figure 3: Equi-probability lines for the four different focus conditions depending on the accent command amplitudes of the two peaks. The stars denote the approximate values for the produced averages of the four options.

on the latter word of two-accent utterance. However, the first accented word receives the perception of narrow focus with only a 1.5 semitone raise – given that the second word is completely unaccented. The F_0 rise of the narrowly focused item must be increased accordingly when the F_0 of other item rises. Only when both rises reach a value of approximately 3.6 semitones, are both target words are perceived as focused simultaneously. Stimuli with rises closest to the means given for the four conditions above were generally identified as belonging to the intended category, though the condition "broad", apparently the default choice, covers a large triangular region in the two-dimensional accent space (see Figure 3).

With respect to statistical analysis, the accent amplitude parameter shows a clear influence on subjects' responses. Analyses of variance with the mean responses (rounded to the closest category) as grouping variable show an F value of 12.334 for Aa1 and of 44.867 for Aa2, both at a significance level of $p < 0.0001$. Since the accent command amplitudes correlate with the other scales in our experiment we can expect to have similar levels of significance for the semitone and Hertz values.

4 Conclusion

It is interesting to note that the Fujisaki parameters correlate to a degree with perception. This may well be due to the fact that the phrase component in our study was not varied. For varying the phrase component would lead to variation between the rises and falls of the accents. In fact, there is some evidence ([9]) that the perception of the last accent of the utterance is conditioned by its degree of fall rather than the rise.

The phrase component and, thus, declination was not varied in the experiment. Therefore, we can only draw tentative conclusions concerning a possible abstract baseline as first reported by Pierrehumbert [1]. However, the presence of a declining baseline in Finnish has been well attested (see, for instance [10]). The fact that in our experiment, the perception of the category "broad" requires the second peak in the utterance to be lower than the first one, gives some evidence to the hypothesis that Finnish listeners and speakers normalize for the baseline declination.

References

[1] Janet Pierrehumbert. The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, 66:363–369, 1979.

[2] Carlos Gussenhoven and T. Rietveld. Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonecits*, 16:355–369, 1988.

[3] Jacques Terken. Fundamental frequency and perceived prominence of accented syllables II: Non-final syllables. *Journal of the Acoustical Society of America*, 95:3662–3665, 1994.

[4] D. Robert Ladd, Jo Verhoeven, and Karen Jacobs. Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects. *Journal of Phonetics*, 22, 1994.

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

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

[7] Hansjörg Mixdorff, Martti Vainio, Stefan Werner, and Juhani Järvikivi. The Manifestation of Linguistic Information in Prosodic Features of Finnish. In Bernard Bel and Isabelle Marlien, editors, *Proceedings of Prosody 2002*, pages 515–518, 2002.

[8] Jacques Terken and Dik Hermes. The perception of prosodic prominence. In Merle Horne, editor, *Prosody: Theory and Experiment*, pages 89–127. Kluwer Academic Publishers, 2000.

[9] Martti Vainio, Juhani Järvikivi, and Stefan Werner. Perception of Focus in Finnish: Influence of Tonal Features and Word Order. Unpublished work in preparation.

[10] P. Hirvonen. *Finnish and English Communicative Intonation*. University of Turku, 1970.