Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Phonetics 31 (2003) 113–138

Journal of Phonetics

www.elsevier.com/locate/phonetics

Durational and tonal correlates of accent in Finnish

Kari Suomi*, Juhani Toivanen, Riikka Ylitalo

Department of Finnish and Saami and Logopedics, Oulu University, P.O. Box 1000, 90014, Finland

Received 22 March 2001; received in revised form 23 May 2002; accepted 16 October 2002

Abstract

The paper reports the first study that explicitly distinguishes the phonetic correlates of sentence accents from those of word stress in Finnish (stress acted as the baseline whose correlates as against of stress were not investigated). Sentences were constructed that successfully elicited no accent, moderate accent, and strong accent on target words. The three degrees of prominence were clearly differentiated phonetically. Mere word stress was not signalled tonally, while accents were signalled mainly tonally. Strong accent involved longer segmental durations than the other degrees of prominence. Timing and extent of the accentual tonal movements were highly uniform across different word structures, and not invariably tied to the initial (stressed) syllables. The data were consistent with a view that the timing of f_0 movements was dependent on the moraic structure of target words. This finding seems to be connected with the typologically rare combination of stress (or rhythmic) and quantity systems in the language.

1. Introduction

In past empirical studies of Finnish prosody, (word) stress and (sentence) accent have not been systematically distinguished; instead, what are actually utterance-level prominences have usually been equated with stress (as has also often been the case with more thoroughly investigated languages). Consequently, this is probably the first attempt to dissociate the realization of stress from that of accent in the language. Different views of the phonetic nature of stress have been presented. For some researchers mere stress, i.e., stress not accompanied by accent, has no phonetic reality at all. Thus Lehiste, for example, claims that "word-level stress is in a very real sense an abstract quality: a potential for being stressed. Word-level stress is the capacity of a syllable to receive sentence-stress when the word is realized as part of the sentence" (Lehiste, 1970, p. 150). For others, stress does have a phonetic reality of its own. For example, Huss (1978)

*Corresponding author.
E-mail address: kari.suomi@oulu.fi (K. Suomi).

0095-4470/03/$ - see front matter
PII: S0095-4470(02)00074-8

demonstrated that, in English, syllables carrying stress differed reliably from unstressed ones in terms of durations and intensities, and more recently Sluijter and van Heuven (1996) showed that, in Dutch, durational and intensity (spectral tilt) differences exist between stressed and unstressed syllables, independently of accent. Findings such as these are consistent with the suggestions in recent prosodic models of e.g., English (Cruttenden, 1997) and Swedish (Bruce, 1998) that the phonetic correlates of stress in these languages are complex but nontonal, while accents in turn are primarily signalled by tonal cues. There are at least two circumstances, apart from genetic differences, that warn against an a priori assumption that the situation in Finnish is similar to that in the more systematically studied languages just mentioned. First, Finnish has fixed word-initial stress, against the moving stress in the Germanic languages. Since (primary) stress is always associated with word onsets, it is conceivable that its realization is different from that in languages in which stress cannot possibly serve a demarcative function. Second, Finnish is a full-fledged quantity language with a quantity opposition in both consonants and vowels, and the oppositions are available irrespective of stress (and not only in stressed syllables, like the opposition between short and long vowels in the languages mentioned above). In Finnish there are contrasting word structures like CVCV, CVVCV, CVCCV, CVVCCV, CVCVV, CVCCVV, etc. in which sequences of adjacent V's and C's may consist of identical phonemes. It is not impossible that in such a language, in which relative segment durations serve the purpose of distinguishing words (both lexemes and word forms) to such a high extent, duration cannot signal prosodic distinctions as freely as in languages in which segmental quantity is less pervasive. Thus, both K'ekchi and Cakchiquel (two Mayan languages) have fixed word-final stress but while K'ekchi has phonemic vowel length, Cakchiquel does not, and Berinstein (1979) observed that Cakchiquel speakers used variations in duration as a perceptual cue for stress whereas the K'ekchi speakers did not (Berinstein did not distinguish between word stress and sentence accent, and seems in fact to have investigated the latter).

Finnish appears to be typologically rare in its specific combination of stress (or rhythmic) system and quantity system; the following account is based on Hayes (1995). Finnish exhibits a bounded stress system: in such systems stresses fall within a particular distance of a boundary or another stress. Finnish is also a *syllabic trochee* language: in such languages feet consist of two syllables, where the first syllable is metrically strong irrespective of its weight (and hence, of the number of morae it contains). Accordingly, primary stress in Finnish always falls on the first syllable following a word boundary, secondary stresses usually on following odd syllables. Hayes (1995, p. 101) argues that, "statistically, syllabic trochee languages tend to be languages that have no quantity distinction at all—that is, no vowel length contrast and no phonological rules that distinguish syllable weight. It is plausible to suppose that in such languages, every syllable is monomoraic". Hayes (1995) then notes that 10 languages in his sample, including Finnish (together with no less than four more or less closely related languages, namely Hungarian, Estonian, Mansi and Votic), "apparently go against the general tendency: they have a vowel length distinction (and in some cases, additional evidence of heavy syllables) but a stress system analyzable with syllabic trochees". That is, in 'pure' syllabic trochee languages there are neither syllable quantity (alias syllable weight) distinctions nor segmental quantity distinctions. But if a syllabic trochee language does have a segmental (especially vowel) quantity distinction, it should also exhibit a sensitivity to syllable weight in its assignment of stress (by avoiding stresses on light syllables). Finnish, then, is exceptional among the syllable trochee languages in having a quantity distinction, and in that syllable weight, nevertheless, plays no part in the assignment of primary stress.[1]

Syllabic trochee languages with a quantity system thus usually exhibit sensitivity to syllable weight in their stress assignment. An implication of this circumstance is that, in such languages, a light syllable alone is somehow unable or insufficient to accommodate prominence. There are indications that this may be the case in Finnish: although phonologically it is unquestionable that it is always the first syllable of a word to which stress is assigned, phonetically the prominence may be signalled (at least partly) later in the word. Thus Lehtonen (1970), in whose main experiment the words studied received moderate accentuation, compared, in a separate experiment, the durations so obtained to ones obtained under strong accentuation (but for one speaker only, and using a very limited material), and concluded that "the domain of stress [in our terminology: accentuation] was shown to be not the first syllable but the first sequence of two moras. For this reason, a Finnish person 'hears' the first short syllable as stressed even when the peaks of intonation, intensity and duration are on the second syllable of the word" (Lehtonen, 1970, p. 151). Second, as Lehtonen also points out, it is well known that foreigners often report hearing that stress in words of the structure (C)V.CVX (in which the indicated location of the syllable boundary is uncontroversial) falls on the second, and not the initial syllable (several personal communications to the first author; see also Collinder, 1937; Sadeniemi, 1949). A very conspicuous consequence of the moraic structure of the word-initial syllable concerns the duration of the second-syllable single vowel: if the first syllable is monomoraic (or light), as in word structures (C)V.CVX, then the duration of the second-syllable single vowel—which constitutes the second mora of the word—is usually very long, much longer than that of the first-syllable single vowel ("usually", because there is dialectal variation), but it is very short when the first syllable is polymoraic (or heavy); for results of relevant measurements see Section 2.3.1.

2. Experiment 1

To investigate the phonetic correlates of stress and accent in Finnish an experiment was conducted in which the degree of prominence of the target words was carefully controlled. The effect of moraic structure on how prominence is realized was controlled by choosing disyllabic target words that jointly represented three moraically different structures.

2.1. Methods

2.1.1. Subjects

Talkers were 10 volunteers, female first-year students of logopedics at Oulu University aged around 20 years without any known hearing or speaking disorders, all speaking the same northern variety of Finnish.

[1] But syllable weight does play a part in the assignment of secondary stress. The rule is: Secondary stresses usually fall on odd syllables but not on the word-final one, except that a heavy final syllable may receive secondary stress if the preceding syllable is light. The rule says "usually" because in some cases secondary stresses are morphologically conditioned, and because secondary stresses are not always fixed.

2.1.2. Materials

Three types of target words were used, viz. CV.CV, CV.CVV, and CVV.CV (hereafter, the locations of the syllable boundaries will not be indicated). The first two types had a monomoraic first syllable, the third a bimoraic one. The structure CVCVV was included partly because we wanted to see how the opposition V vs. VV is realized after a light initial syllable in this particular variety of Finnish; this aspect of the results will not be commented on below. The target words, numbering 11 of type CVCV, 8 of type CVCVV and 8 of type CVVCV, were all common nouns. An example of an ideal triplet of words representing these three structures is *lama*, *lamaa*, *laama*, a triplet in which only the quantity relations are different, but the phonetic segments (apart from durational differences signalling the quantity relations) are identical across the structures. Because of gaps in the vocabulary, and our wish to avoid segments that are notoriously difficult to segment consistently (e.g., /r/ that has variable realizations), only five such ideal triplets of words (representing common nouns) could be found. To increase the number of target words, phonetically similar word pairs (rather than triplets) had to be included, with each pair representing the CVCV structure and one of the other structures. Examples are *kato* (CVCV) and *kaato* (CVVCV), and *kota* (CVCV) and *kotaa* (CVCVV). Details aside, facts of Finnish morphology and the existence of accidental gaps in the vocabulary strictly limited the choice of words; for the same reasons, it was not possible to have equal numbers of words in the three structural types. The consequences of the segmental imbalance among the target word groups on the results will be commented on below. The words used are listed in the appendix.

For each target word a triplet of short sentences was constructed. In the first sentence type, e.g., “Sanoin että kato PAHENTAA tilannetta, en sanonut että kato PARANTAA sitä” (‘I said that dearth WORSENS the situation, I didn’t say that dearth IMPROVES it’), a contrastive accent was indicated on the verb following the target (*kato* ‘dearth’), and the target was mentioned twice. It was always the second occurrence of the target that was measured, in order to obtain as deaccented a version as possible. Below, we will refer to this condition as Word Stress. In the second sentence type, e.g. “Sanoin että kato pitää tilanteen edelleen KIREÄNÄ” (‘I said that dearth continues to keep the situation TENSE’), the target occurred early in the sentence, and an emphatic word followed at the end of the sentence. This condition was intended to elicit a moderately accented version of the target, and it will be referred to as Moderate Accent below. In the third sentence type, e.g., “Sanoin että KATO pelotti, en sanonut että KATTO pelotti” (‘I said that DEARTH frightened (me), I didn’t say that the CEILING frightened (me)’), the target itself carried a contrastive accent, and below this condition will be referred to as Strong Accent. The target noun was always either the uninflected (i.e., nominative singular) form functioning as the subject of the sentence (as in the example triplet just given), or, in passive constructions, the partitive singular form functioning as the object; for phonotactic and syntactic reasons all CVCVV words were in fact of this form. The target word was always preceded by the conjunction *että* (‘that’) and, in a given triplet, followed by the same plosive. It will be seen from the examples that the carrier sentences differed in length and in other respects, and in particular the Word Stress version of each target occurred later in its carrier sentence than did the other versions in theirs.

There were 27 (target words) $\times$ 3 (sentence types or target versions) = 81 different sentences. These were randomized, with the restriction that a given word was not allowed to occur in two consecutive sentences. To the beginning of the final randomized list of sentences, five practice sentences were added.

2.1.3. Recording

The written sentences were presented to the talker on a computer screen placed outside the monitoring window of the recording studio, and the talker controlled the speed of presentation by clicking the mouse. After saying a sentence, the talker had to wait 2 s before she could click for the next sentence. Talkers were instructed to speak the sentences in a way they deemed natural, and to highlight the capitalized words. They were also instructed to repeat a sentence if they detected an error (otherwise no repetitions were elicited). We deliberately abstained from giving more detailed instructions, and from on-line monitoring of the productions, in order not to affect the talkers' natural manner of speaking. For example, we gave no instructions as to pausing. The talkers' productions were recorded using a high-quality microphone and an MD recorder, and transferred onto hard disk (44.1 kHz, 16 bit).

2.1.4. Measurements

Two of the authors performed all of the acoustic measurements. Durations of the following words/segments were measured for each version of each target word: the preceding conjunction *että*; C_1, V_1, C_2, and V_2 of the target (V_1 and V_2 here and subsequently denoting phonetic vowel segments irrespective of the number of phonemes they represent). Burst portions of voiceless plosives were included in the duration of the consonant; the Finnish voiceless plosives are unaspirated, Suomi (1980) reporting mean VOT values of 10 ms for /p/, 13 ms for /t/, and 21 ms for /k/, pooled across word-initial and medial positions (10 speakers). F_0 was measured (to the nearest tenth of a Hz) at the beginning and end of the syllable preceding the target (but for microprosodic effects see below); at three locations, viz. beginning, middle and end, of the voiced portion of both syllables of the target; and the corresponding three locations of the syllable following the target. Two additional f_0 measurements were made for the phonetically long syllables of the targets, i.e., for the second syllables of CVCV(V) words and the initial syllables of CVVCV words. These latter measurement points were located halfway between the beginning and middle, and middle and end, respectively. When the medial consonant was a voiced one and phonation thus continued without interruption from the first syllable to the second, the end f_0 of the first syllable was measured at the last but one glottal pulse before the syllable boundary and the beginning f_0 of the second syllable was measured at the second glottal pulse after the syllable boundary (in order to avoid measuring adjacent glottal pulses). After all (initial or medial) voiceless plosives an attempt was made to avoid microprosodic effects by not placing the measurement point immediately after the onset of phonation; this also applies to the syllable preceding the target word, the second syllable of the conjunction *että*. When the f_0 measurements were begun, the analytical adequacy of the intended set of measurement points was continuously evaluated. It was estimated whether the set of measurement points as defined would be sufficient to capture the essential properties of the raw f_0 contours, especially the minimum and peak values, or whether even more measurement points would be needed. It turned out that three measurement points for short syllables and five for long ones were sufficient in this respect. In particular, it turned out that the minimum and maximum f_0 values always occurred at or very near a measurement point. Thus although minimum and maximum values were not explicitly identified when the measurements were taken, they were nevertheless very closely recorded (a pilot study had taught us that three measurement points for all kinds of syllables would not be sufficient in this respect). Altogether, then, the measured beginning, middle and end points are somewhat relative (and recall that they represent the voiced portions of the syllables).

One of the authors auditorily evaluated the prominences of the 810 word tokens in their carrier sentences, unaware of the results of the acoustic measurements, using a three-way scaling: not accented, moderately accented, and strongly accented. The evaluator, who is very experienced in such a task, was inevitably aware of the intended prominence of each token, but made every effort to perform the scaling only on the basis of how each individual token had been actually produced, listening to each token many times both in and out of context; it was of course very much in the interest of the evaluator that the tripartite division would be perceptually as accurate as possible. As it happened, the outcome of this evaluation was highly consistent with the scaling that was attempted in constructing the three sentence types (not accented corresponding to Word Stress, etc.); we had, in fact, been sceptical about whether it would be possible to rely on a three-way scaling consistently. Twenty-one (2.6%) of the spoken tokens were rejected because of a discrepancy between the intended and the evaluated degree of prominence, because of disfluencies (including pauses which, if occurring in the immediate vicinity of the target word, invariably occurred *after* the target word), or because of obvious mispronunciations.

2.2. Results

Separately for each dependent variable, two types of repeated measures ANOVAs were performed, one with subjects (i.e., speakers) as the random factor ($F1$), and one with items (i.e., words) as the random factor ($F2$). In the subjects analyses, both Prominence (Strong Accent, Moderate Accent, Word Stress) and Structure (CVCV, CVVCV, CVCVV) were treated as within-subject factors—as all speakers produced all target words with each degree of prominence—whereas in the items analysis Prominence was treated as a within-subject (or within-item) factor and Structure as a between-subject (or between-item) factor—since all target words occurred in three degrees of prominence, but groups of target words inevitably differed with respect to their structure. Pairwise comparisons were similarly performed separately for subjects ($t1$) and for items ($t2$); tests of significance in t-tests were always two-tailed. Occasionally, we used t-tests in the subjects analysis but one-way ANOVAs in the items analysis (because in the latter, group sizes were not identical). In all such tests, we regard as reliable only differences of means that were significant in both analyses.

2.2.1. Durations

The results of the durational measurements are shown in Table 1. Prominence had a reliable effect on the duration of *että* [$F1(1,9)=42.46$, $p<0.001$; $F2(1,24)=86.68$, $p<0.001$], while Structure had no effect [$F1(1,9)=1.11$, ns; $F2(2,24)<0$] and there was no interaction between Prominence and Structure [$F1(1,9)<0$; $F2(2,24)<0$]. Pairwise comparisons indicated that Strong Accent induced longer duration of the conjunction than did Moderate Accent [$t1=8.78$, $p<0.001$, df$=29$; $t2=11.48$, $p<0.001$, df$=26$] and Word Stress [$t1=9.67$, $p<0.001$, df$=29$; $t2=9.57$, $p<0.001$, df$=26$], and that Moderate Accent induced longer duration than Word Stress [$t1=2.36$, $p<0.05$, df$=29$; $t2=2.25$, $p<0.05$, df$=26$]. Thus, the duration of the preceding *että*

Table 1
Mean durations (in ms) of the preceding *että* and of the segments of the target word types in the three degrees of prominence

Measured unit	Word structure		
	CVCV	CVVCV	CVCVV
että			
Strong Accent	231	229	233
Moderate Accent	208	212	210
Word Stress	201	208	205
C_1			
Strong Accent	108	110	119
Moderate Accent	82	85	87
Word Stress	82	87	90
V_1			
Strong Accent	78	198	78
Moderate Accent	58	143	64
Word Stress	56	144	62
C_2			
Strong Accent	81	63	87
Moderate Accent	68	56	72
Word Stress	72	61	78
V_2			
Strong Accent	137	57	199
Moderate Accent	104	50	144
Word Stress	103	53	154

increased incrementally as a function of increasing prominence on the following word. Prominence had a reliable effect on the duration of C_1 [$F1(1,9) = 154.82$, $p < 0.001$; $F2(1,24) = 331.65$, $p < 0.001$], Structure failed to have a significant effect in the items analysis [$F1(1,9) = 1.63$, $p < 0.01$]; $F2(2,24) < 0$] and there was no interaction [$F1(1,9) > 0$; $F2(2,24) = 1.62$, ns]. Pairwise comparisons indicated that Strong Accent induced longer duration of C_1 than did Moderate Accent [$t1 = 17.80$, $p < 0.001$, df = 29; $t2 = 15.97$, $p < 0.001$, df = 26] and Word Stress [$t1 = 12.46$, $p < 0.001$, df = 29; $t2 = 18.05$, $p < 0.001$, df = 26], while there was no difference between the latter two versions [$t1 = 1.14$, ns, df = 29; $t2 = 1.10$, ns, df = 26].

All effects on V_1 were fully significant: Prominence [$F1(1,9) = 81.03$, $p < 0.001$; $F2(1,24) = 334.57$, $p < 0.001$], Structure [$F1(1,9) = 7.39$, $p < 0.05$; $F2(2,24) = 377.50$, $p < 0.001$], and the interaction [$F1(1,9) = 19.41$, $p < 0.01$; $F2(2,24) = 47.89$, $p < 0.001$]. Pairwise comparisons (subjects analysis) and one-way ANOVAs (items analysis) were performed to compare the three word structures in each degree of prominence. In Strong Accent, the pairwise comparisons indicated that CVVCV words differed from both CVCV words [$t1 = 18.88$, $p < 0.001$, df = 9] and from CVCVV words [$t1 = 17.71$, $p < 0.001$, df = 9], and that there was no difference between the latter two ($t < 0$). In the one-way ANOVA the effect of Structure was fully significant [$F2(2,24) = 369.37$, $p < 0.001$], and Scheffe post hoc tests indicated that CVVCV words differed from both CVCV words [$p < 0.001$] and CVCVV words [$p < 0.001$], and that there was no difference between the latter two (ns). The same pattern was observed in the other two degrees of prominence. In Moderate accent the pairwise comparisons yielded [$t1 = 20.54$, $p < 0.001$, df = 9] for the CVCV–CVVCV difference, [$t1 = 2.97$, $p < 0.05$, df = 9] for the CVCV–CVCVV difference, and [$t1 = 18.92$, $p < 0.001$, df = 9] for the CVVCV–CVCVV difference. In items analysis, the effect of Structure was again fully significant [$F2(2,24) = 198.46$, $p < 0.001$], and Scheffe post hoc results were as in Strong Accent (that is, there was no difference between CVCV and CVCVV in the items analysis). In Word Stress, finally, subjects analysis produced [$t1 = 16.37$, $p < 0.001$, df = 9] for CVCV–CVVCV, [$t1 = 4.06$, $p < 0.01$, df = 9] for CVCV–CVCVV, and [$t1 = 14.67$, $p < 0.001$, df = 9] for CVVCV–CVCVV, while items analysis produced [$F2(2,24) = 334.92$, $p < 0.001$] and Scheffe post hoc results were as in Strong Accent (that is, there was again no difference between CVCV and CVCVV in the items analysis). In all degrees of prominence, then, CVVCV differed fully reliably from both CVCV and CVCVV, but the latter two structures failed to be reliably different in the items analyses. Thus, quite expectedly, the categorical V vs. VV opposition in the first syllable was amply maintained in all degrees of prominence. Notice that the duration of the first-syllable VV is about 2.5 times longer than that of the first-syllable V in each degree of prominence, which is interpretable as a lengthening of the word's second mora (because without such a lengthening, one would expect VV to have only twice the duration of V).

Turning now to the effects of Prominence on the duration of V_1, pairwise comparisons indicated that in structure CVCV Strong Accent differed from both Moderate Accent [$t1 = 6.32$, $p < 0.001$, df = 9; $t2 = 9.77$, $p < 0.001$, df = 10] and Word Stress [$t1 = 9.13$, $p < 0.001$, df = 9; $t2 = 8.74$, $p < 0.001$, df = 10], and that there was no difference between the latter two [$t1 = 1.75$, ns, df = 9; $t2 = 2.68$, $p < 0.05$, df = 10]. In structure CVVCV too Strong Accent differed from both Moderate Accent [$t1 = 6.31$, $p < 0.001$, df = 9; $t2 = 16.16$, $p < 0.001$, df = 7] and Word Stress [$t1 = 7.45$, $p < 0.001$, df = 9; $t2 = 13.79$, $p < 0.001$, df = 7], and again there was no difference between the latter two [$t1 < 0$; $t2 < 0$]. In structure CVCVV, finally, Strong Accent again differed from both Moderate Accent [$t1 = 4.68$, $p < 0.001$, df = 9; $t2 = 13.13$, $p < 0.001$, df = 7] and Word Stress [$t1 = 8.60$, $p < 0.001$, df = 9; $t2 = 9.08$, $p < 0.001$, df = 7], and once more there was no difference between the latter two [$t1 = 1.06$, ns, df = 9; $t2 = 1.43$, ns, df = 7]. That is, V_1 was consistently longer in Strong Accent than in Moderate Accent and Word Stress, and there was no difference between the latter two.

For C_2 there was a significant effect of Prominence [$F1(1,9) = 16.00$, $p < 0.01$; $F2(1,24) = 19.01$, $p < 0.001$], an effect of Structure [$F1(1,9) = 40.80$, $p < 0.001$; $F2(2,24) = 4.75$, $p < 0.05$], and no interaction [$F1(1,9) < 0$; $F2(2,24) = 2.89$, ns]. Differences among the word structures within each degree of prominence were again examined using pairwise comparisons (subjects analysis) and one-way ANOVAs (items analysis). In Strong Accent the pairwise comparisons indicated reliable differences among all three structures: for CVCV–CVVCV [$t1 = 13.55$, $p < 0.001$, df = 9], for CVCV–CVCVV [$t1 = 4.74$, $p < 0.01$, df = 9], for CVVCV–CVCVV [$t1 = 14.34$, $p < 0.001$, df = 9]. In items analysis the effect of Structure was significant [$F2(2,24) = 4.88$, $p < 0.05$], but Scheffe post hoc tests indicated that only the CVVCV–CVCVV difference reached significance [$p < 0.05$]. In Moderate Accent subjects analysis produced [$t1 = 5.19$, $p = 0.001$, df = 9] for CVCV–CVVCV, [$t1 = 2.83$, $p < 0.05$, df = 9] for CVCV–CVCVV, and [$t1 = 6.47$, $p < 0.001$, df = 9] for CVVCV–CVCVV. In items analysis the effect of Structure was again significant [$F2(2,24) = 5.29$, $p < 0.05$], and Scheffe post hoc tests again indicated that only the CVVCV–CVCVV difference reached significance [$p < 0.05$]. In Word Stress, finally, subjects analysis produced [$t1 = 6.74$, $p < 0.001$, df = 9] for CVCV–CVVCV, [$t1 = 4.44$, $p < 0.01$, df = 9] for

CVCV–CVCVV, and [$t1 = 9.84$, $p < 0.001$, df = 9] for CVVCV–CVCVV, but in the items analysis Structure very narrowly failed to reach significance [$F2(2,24) = 3.25$, $p = 0.056$]. Thus the duration of C2 was longer in CVCVV than in CVVCV in both degrees of accentuation (and in Word Stress the numerical difference of the means was in the same direction). It is known from Lehtonen's (1970) study that a double vowel lengthens the duration of the preceding consonant in at least the word's first, second and third syllable: in all relevant comparisons examined by Lehtonen (CV.CV vs. CV.CVV, CV.CVC vs. CV.CVVC, CVV.CV vs. CVV.CVV, CV.CVV.CV vs. CV.CVV.CVV, and CVV.CVV.CV vs. CVV.CVV.CVV), he found that a single consonant was fully reliably longer before a double vowel than before a single one, the mean difference computed across the comparisons just mentioned being 13.2 ms.

As for the effects of Prominence, pairwise comparisons indicated that in CVCV words the duration of C_2 was reliably longer in Strong Accent than in Moderate Accent [$t1 = 8.62$, $p < 0.001$, df = 9; $t2 = 5.22$, $p < 0.001$, df = 10] and in Word Stress [$t1 = 3.96$, $p < 0.01$, df = 9; $t2 = 4.15$, $p < 0.01$, df = 10], and that there was no difference between the latter two [$t1 = 1.87$, ns, df = 9; $t2 = 2.26$, $p < 0.05$, df = 10]. In CVVCV words C2 was longer in Strong Accent than in Moderate Accent [$t1 = 3.09$, $p < 0.05$, df = 9; $t2 = 2.56$, $p < 0.05$, df = 7], but there was no difference between Strong Accent and Word Stress [$t1 = 1.21$, ns, df = 9; $t2 < 0$] nor between the latter and Moderate Accent [$t1$ 5.34, $p < 0.001$, df = 9; $t2 = 1.85$, ns, df = 7]. In CVCVV, finally, C_2 was longer in Strong Accent than in Moderate Accent [$t1 = 8.35$, $p < 0.001$, df = 9; $t2 = 4.13$, $p < 0.01$, df = 7] and in Word Stress [$t1 = 4.85$, $p < 0.001$, df = 9; $t2 = 3.11$, $p < 0.05$, df = 7], and C_2 was shorter in Moderate Accent than in Word Stress [$t1 = 3.98$, $p < 0.01$, df = 9; $t2 = 2.50$, $p < 0.05$, df = 7]. In this complex pattern, the only reliable difference of means that was found for all three word structures, was that C_2 had a longer duration in Strong Accent than in one or two of the other degrees of prominence. Otherwise, C_2 had longer duration in Word Stress than in Moderate Accent in CVCVV words (and there were very small numerical differences in the same direction in CVCV and CVVCV words). We cannot think of any explanation for this last observation.

For V_2 all effects were fully significant: Prominence [$F1(1,9) = 96.74$, $p < 0.001$; $F2(1,24) = 282.61$, $p < 0.001$], Structure [$F1(1,9) = 135.79$, $p < 0.001$; $F2(2,24) = 375.09$, $p < 0.001$], and the interaction [$F1(1,9) = 11.38$, $p < 0.001$; $F2(2,24) = 47.14$, $p < 0.001$]. Looking at Structure first, it was fully and predictably significant in all three degrees of prominence. In Strong Accent pairwise comparisons yielded [$t1 = 15.02$, $p < 0.001$, df = 9] for CVCV–CVVCV, [$t1 = 11.44$, $p < 0.001$, df = 9] for CVCV–CVCVV, and [$t1 = 19.81$, $p < 0.001$, df = 9] for CVVCV–CVCVV, in items analysis the main effect of Structure was significant ([$F2(2,24) = 402.01$, $p < 0.001$]), and Scheffe post hoc tests indicated reliable differences between all three structures (in each comparison, $p < 0.001$). In Moderate Accent, pairwise comparisons yielded [$t1 = 14.87$, $p = 0.001$, df = 9] for CVCV–CVVCV, [$t1 = 7.89$, $p < 0.001$, df = 9] for CVCV–CVCVV, and [$t1 = 22.85$, $p < 0.001$, df = 9] for CVVCV–CVCVV, and in items analysis Structure was again significant ([$F2(2,24) = 196.43$, $p < 0.001$]), with the results of Scheffe post hoc tests as in Strong Accent. In Word Stress, finally, differences between the three structures were significant in the subjects analysis ([$t1 = 11.04$, $p < 0.001$, df = 9] for CVCV–CVVCV, [$t1 = 11.61$, $p < 0.001$, df = 9] for CVCV–CVCVV, and [$t1 = 26.37$, $p < 0.001$, df = 9] CVVCV–CVCVV), as they were in the items analysis: there was a main effect of Structure ([$F2(2,24) = 230.34$, $p < 0.001$]), and the results of Scheffe post hoc tests as in Strong Accent. The durations of the second-syllable single vowels in CVCV and CVVCV words illustrate how dramatically the duration of those vowels depends on whether or not they constitute the word's second mora: V_2 in CVCV is approximately twice as long as V_2 in CVVCV.

Turning to the effects of Prominence on V_2, pairwise comparisons indicated that in structure CVCV V_2 had a fully reliably longer duration in Strong Accent than in Moderate Accent [$t1=7.17$, $p<0.001$, df$=9$; $t2=10.43$, $p<0.001$, df$=10$] and in Word Stress [$t1=9.11$, $p<0.001$, df$=9$; $t2=11.30$, $p<0.001$, df$=10$], and that there was no difference between Moderate Accent and Word Stress [$t1<0$; $t2<0$]. In CVVCV V_2 had a longer duration in Strong Accent than in Moderate Accent [$t1=3.99$, $p<0.01$, df$=9$; $t2=4.69$, $p<0.01$, df$=7$], but there was no difference between Strong Accent and Word Stress [$t1=1.94$, ns, df$=9$; $t2=3.43$, $p<0.05$, df$=7$] nor between the latter and Moderate Accent [$t1=1.82$, ns, df$=9$; $t2=1.17$, ns, df$=7$]. In CVCVV, finally, V_2 had a longer duration in Strong Accent than in Moderate Accent [$t1=8.77$, $p<0.001$, df$=9$; $t2=13.62$, $p<0.001$, df$=7$] and in Word Stress [$t1=10.55$, $p<0.001$, df$=9$; $t2=13.16$, $p<0.001$, df$=7$], but there was no difference between Moderate Accent and Word Stress [$t1=2.56$, $p<0.05$, df$=9$; $t2=1.99$, ns, df$=7$]. Thus in both CVCV and CVCVV words V_2 had a fully reliably longer duration in Strong Accent than in both Moderate Accent and Word Stress, whereas in CVVCV V_2 only had a longer duration in Strong Accent than in Moderate Accent, but the other differences were not significant. Comparing the mean obtained in Strong Accent with the mean of the other two degrees of prominence, the lengthening of V_2 accompanying Strong Accent was 33.5 ms (or 32%) in CVCV, only 5.5 ms (11%) in CVVCV, and 50 ms (34%) in CVCVV. Thus, the duration of V_2 lengthened in Strong Accent (both in absolute and proportional terms) much more when V_2 constituted (CVCV) or contained (CVCVV) the word's second mora than when it did not (CVVCV). In the case of V_1, by contrast, the lengthening accompanying Strong Accent was very similar across the word structures: 21 ms (37%) lengthening in CVCV, 54.5 ms (28%) in CVVCV, and 15 ms (24%) in CVCVV (from the perspective of speech perception, the amount of proportional lengthening is probably more important than the amount of lengthening in absolute terms).

As already mentioned above, the three structural word groups were not completely matched in terms of their segmental phonetic structure. In order to estimate the effects of this lack of balance on the results, control analyses were run involving only those triplets of words that consist of exactly the same phonetic segments in each of the three word structures, i.e., words that differ only in terms of the quantity opposition. There were five such fully matched triplets, e.g., *lama*, *lamaa*, *laama* (see the appendix). From this material, we first computed the means corresponding to those presented in Table 1. The average difference between these new means and those in Table 1 was 3.3% (range: 0–11%). Next, all differences of means in this smaller material were tested statistically as was done for the whole material (performing both subjects and items analyses, where appropriate). The following statistical differences were observed. For the duration of *että*, the numerical differences were in the same direction as in the whole material, but the difference between Moderate Accent (214 ms) and Word Stress (210) was significant only in the items analysis. Since the items were now segmentally fully balanced, we interpret the failure of the subjects analysis to be a consequence of the smaller size of the material, and propose that there is no need to alter the conclusion reached above that the duration of *että* increased incrementally as a function of increasing prominence on the following word. In C_1 there were no statistical differences between the two materials. With respect to V_1 CVCV and CVCVV were now reliably different in Moderate Accent and Word Stress (while in the whole material this difference failed to be significant in the items analysis in each degree of prominence). Since the items analysis turned out to reach significance when the items were segmentally fully controlled, even though the size of the material was smaller, we conclude that all three words structures were in fact reliably distinguished with respect to the duration of V_1. For C_2 it was observed that all three word structures were now reliably different in each degree of prominence (while in the whole material, in each degree of prominence, only the CVVCV − CVCVV difference reached significance also in the items analysis). Again, since the items analysis now reached significance when item structure was strictly controlled, we conclude that all three word structures were reliably distinguished on the basis of C_2, in each degree of prominence. For V_2, finally, there were no statistical differences between the two sets of materials. Notice that, except for the difference concerning the conjunction *että*, all observed statistical differences relate to the distinguishability of the three word structures. Consequently, given our interpretation of the difference concerning *että* as presented above, the two sets of materials produced statistically identical results as far as the differences between the three degrees of prominence are concerned. These latter differences are our main concern in this paper, and while the means presented in Table 1 may contain a small measure of segment-specific influence (due to differences in inherent segmental durations) in addition to the influence of phonological structure in terms of C's and V's, we prefer to present the results based on the whole material (especially since, as will be shown in the next section, there was practically no difference between the two materials in terms of f_0).

To summarize the durational results of Experiment 1, each measured word or segment was always reliably longer in Strong Accent than in at least one of the other degrees of prominence. The difference of means between Moderate Accent and Word Stress was significant in two comparisons, *että* and C_2 in the CVCVV structure (in which C_2 was longer in Word Stress). In both cases, the differences were very small, and no explanation seems to be readily available. The overall generalisation thus seems to be that Strong Accent involved a longer duration of all segments in the sequence *että* + target word than the other two degrees of prominence. It is not possible to determine the exact domain of the lengthening, for at least two reasons. First, the prosodic affiliation of the preceding conjunction is unclear.[2] Second, each target word simultaneously constituted a disyllabic foot. Finally, recall that, within the target word, the difference between Moderate Accent and Word Stress was significant in only one comparison (and then numerically minute), namely the duration of C_2 in CVCVV. Ignoring this exception, the conclusion is that moderate accentuation was not realized durationally.

2.2.2. *Fundamental frequency*

The target words were always preceded by the conjunction *että*, but as already noted above, the carrier sentences inevitably differed among themselves in other respects, and in particular the Word Stress version of each target occurred later in its carrier sentence than did the other versions in theirs. Therefore, it seemed likely that overall sentence intonation (or declination) would have

[2] In traditional syntactic descriptions based on written language, the conjunction is analyzed to be the first element of the clause that follows the conjunction, and normative rules of punctuation reiterate this analysis by demanding a comma before the conjunction. On the other hand, it is possible that prosodically the conjunction is the last element of the *preceding* unit (for example, it seems unquestionable that pauses in spontaneous speech mostly occur after rather than before the conjuntion). At the moment, however, we are unable to decide on this issue, although the results of Experiment II do seem more consistent with the latter alternative.

an effect on the absolute starting levels of f_0 in the different degrees of prominence, and this was in fact the main reason why f_0 during *että* was measured. An analysis of the f_0 values measured at the end of *että* confirmed the expectation. The effect of Prominence was highly significant [$F1(1,9) = 23.22$, $p < 0.001$; $F2(1,24) = 225.65$, $p < 0.001$], while there was no effect of Structure [$F1(1,9) < 0$; $F2(2,24) < 0$] and no interaction of Prominence and Structure [$F1(1,9) < 0$; $F2(2,24) < 0$]. Since our purpose here was not to investigate effects of global sentence intonation but local f_0 effects of prominence on individual words, it seemed appropriate to normalize the absolute f_0 values relative to the immediately preceding, syntactically and lexically identical context. Consequently, the f_0 values presented below are, for each degree of prominence, values relative to the f_0 at the end of the preceding *että* for that degree of prominence (205 Hz for Strong Accent, 210 Hz for Moderate Accent, and 192 Hz for Word Stress). That is, each normalized f_0 value presented below was computed by subtracting the mean value of *että*, as computed across the degree of prominence in question, from the raw f_0 value: thus for example, a positive value indicates that the raw value was so many Hz higher than that at the end of *että*. A *post hoc* motivation for this decision is that, as will be seen below, the f_0 movements, in each degree of prominence, were highly uniform across the three word structures.

The effects of prominence on the course of f_0 in the CVCV, CVVCV and CVCVV words are shown graphically in Figs. 1, 2 and 3, respectively, and numerically in Table 2. The highly variable durations of the syllables following the target words were not measured, but in the figures their three f_0 measurement points have been given an arbitrary temporal distance of 80 ms. It can be seen that the three degrees of prominence were well differentiated in each word structure, and that

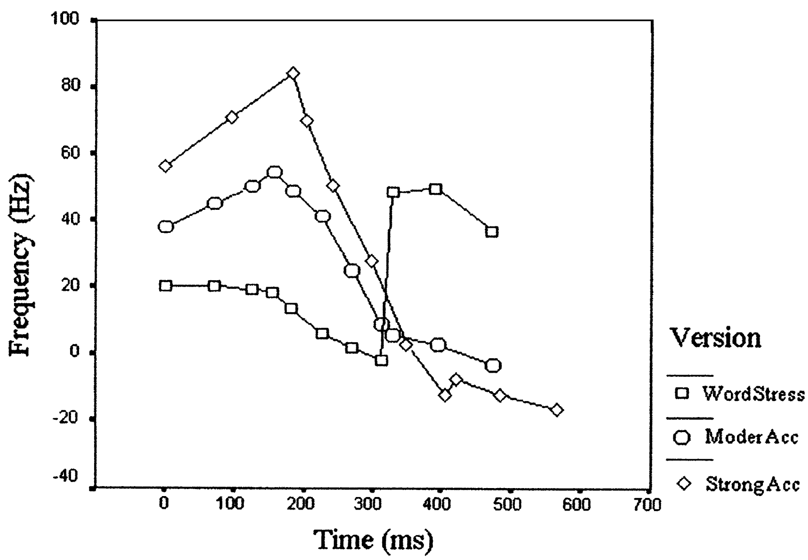

Fig. 1. Mean f_0 values of the CVCV type words in the three degrees of prominence, at the measurement points employed (for more explanation, see the text). For each degree of prominence, the means (in Hz) are relative to the mean f_0 measured at the end of the preceding *että*. Zero milliseconds on the time scale indicates word onset. The syllable boundary is located between the third and the fourth measurement point. The last three measurement points refer to the initial syllable of the word following the target.

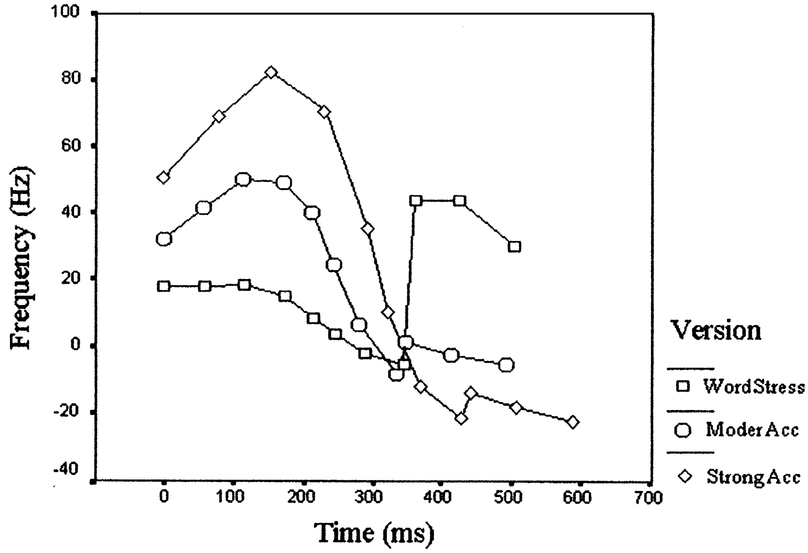

Fig. 2. Mean f_0 values of the CVVCV type words in the three degrees of prominence, at the measurement points employed (for more explanation, see the text). For each degree of prominence, the means (in Hz) are relative to the mean f_0 measured at the end of the preceding *että*. Zero milliseconds on the time scale indicates word onset. The syllable boundary is located between the fifth and the sixth measurement point. The last three measurement points refer to the initial syllable of the word following the target.

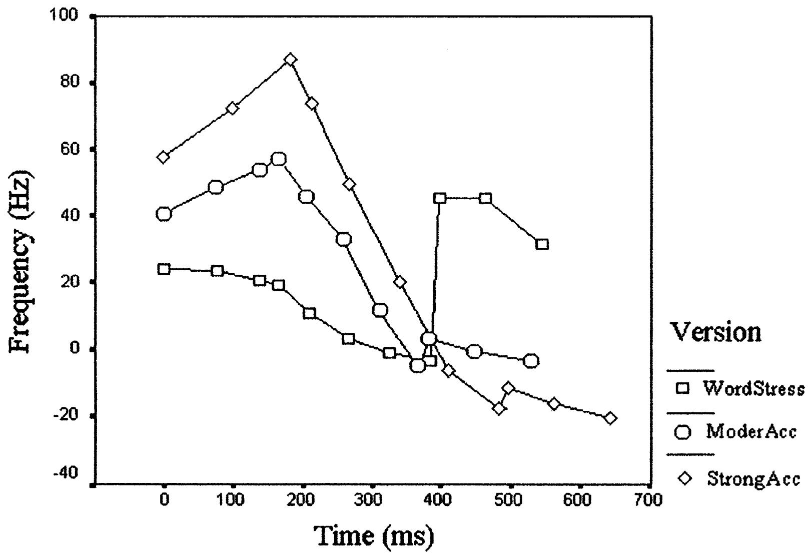

Fig. 3. Mean f_0 values of the CVCVV type words in the three degrees of prominence, at the measurement points employed (for more explanation, see the text). For each degree of prominence, the means (in Hz) are relative to the mean f_0 measured at the end of the preceding *että*. Zero milliseconds on the time scale indicates word onset. The syllable boundary is located between the third and the fourth measurement point. The last three measurement points refer to the initial syllable of the word following the target.

Table 2
Mean fundamental frequencies during the target word types in the three degrees of prominence, at the measurement points employed (for more explanation, see the text). For each degree of prominence, the means (in Hz) are relative to the mean f_0 measured at the end of the preceding *että*

Measurement point	Word structure		
	CVCV	CVVCV	CVCVV
1/1			
Strong Accent	56.3	52.0	57.5
Moderate Accent	37.9	33.4	40.2
Word Stress	20.5	18.7	23.7
1/2			
Strong Accent		70.0	
Moderate Accent		42.5	
Word Stress		18.9	
1/3			
Strong Accent	71.1	83.4	72.3
Moderate Accent	45.3	51.2	48.5
Word Stress	20.6	19.2	23.0
1/4			
Strong Accent		72.1	
Moderate Accent		50.4	
Word Stress		16.0	
1/5			
Strong Accent	84.1	36.7	86.8
Moderate Accent	50.6	41.3	53.5
Word Stress	19.8	9.7	20.2
2/1			
Strong Accent	69.8	11.5	73.4
Moderate Accent	55.2	26.0	57.0
Word Stress	18.8	5.0	19.1
2/2			
Strong Accent	50.9		49.6
Moderate Accent	49.1		45.8
Word Stress	13.3		10.1
2/3			
Strong Accent	28.1	−10.8	19.8
Moderate Accent	41.3	7.7	32.7
Word Stress	6.5	−0.5	2.6
2/4			
Strong Accent	3.2		−6.9
Moderate Accent	25.5		11.2
Word Stress	2.0		−1.5

Table 2 (*continued*)

Measurement point	Word structure		
	CVCV	CVVCV	CVCVV
2/5			
Strong Accent	−12.3	−20.2	−18.4
Moderate Accent	9.1	−7.3	−5.5
Word Stress	−1.4	−3.9	−3.7

the tonal movement was larger in Strong Accent than in Moderate Accent, as expected. The peak f_0 value in Strong Accent always occurred at the third measurement point from word onset. Note carefully that for CVCV and CVCVV words this point corresponds to the end of the first syllable V, for CVVCV words the middle of the first syllable VV. In other words, the peak always occurred near the end of the word's first mora. On the whole the timing, in real time, of the f_0 movements across the three word structures was highly uniform. This is shown for Strong Accent in Fig. 4. As just pointed out, this uniformity does not respect syllabic structure: in CVCV and CVCVV words the rise and the peak occurred on the first syllable and the fall on the second one, whereas in CVVCV words the rise, the peak and much of the fall occurred on the first syllable (see Fig. 2 in which the fifth measurement point from word onset represents the end of the first syllable in CVVCV words).

Figs. 1–4 strongly suggest that, irrespective of word structure, a given degree of accentuation essentially consisted of a steady rise of f_0 to its peak value, followed by a fall (and that Word Stress involved no f_0 peak, and very little f_0 movement at all). In the following, therefore, the results are discussed in terms of mean f_0 values at three measurement points that jointly define the rise–fall pattern of the accents: the values at the beginning of the rise (and target word), at the peak, and at the end of the fall (and the target word). In the analysis of the means at word onset, Prominence had a significant effect [$F1(1,9) = 85.20$, $p < 0.001$; $F2(1,24) = 61.34$, $p < 0.001$], Structure had no effect [$F1(1,9) = 3.33$, ns; $F2(2,24) < 0$], and there was no interaction of Prominence and Structure [$F1(1,9) < 0$; $F2(2,24) < 0$]. Pairwise comparisons showed that Strong Accent differed significantly from Moderate Accent [$t1 = 7.86$, $p < 0.001$, df = 29; $t2 = 6.64$, $p < 0.001$, df = 26] and from Word Stress [$t1 = 14.78$, $p < 0.001$, df = 29; $t2 = 8.28$, $p < 0.001$, df = 26], and that Moderate Accent differed significantly from Word Stress [$t1 = 9.52$, $p < 0.001$, df = 29; $t2 = 8.33$, $p < 0.001$, df = 26]. That is, Strong Accent involved the reliably highest initial f_0 value, and Word Stress the lowest (see Table 2).

The analysis then focused on whether the numerical turning points apparent in Figs. 1–3, i.e., the points at which the mean Hz value is lower than the mean at the previous point at which f_0 reached its peak value, also represent statistically significant changes between consecutive measurement points. Of course, this test makes sense only in the accented versions (as there was no f_0 peak in the Word Stress versions). Looking first at Strong Accent, in CVCV words (Fig. 1) there is an apparent turning point after the third measurement point (end of the first syllable), and pairwise comparisons were made between the means obtained at that point and the next one: $t1 = 2.46$, $p < 0.05$, df = 9; $t2 = 3.17$, $p < 0.01$, df = 10. In CVCVV (Fig. 3) the corresponding comparison produced the same result: $t1 = 2.48$, $p < 0.05$, df = 9; $t2 = 3.65$, $p < 0.01$, df = 7. In

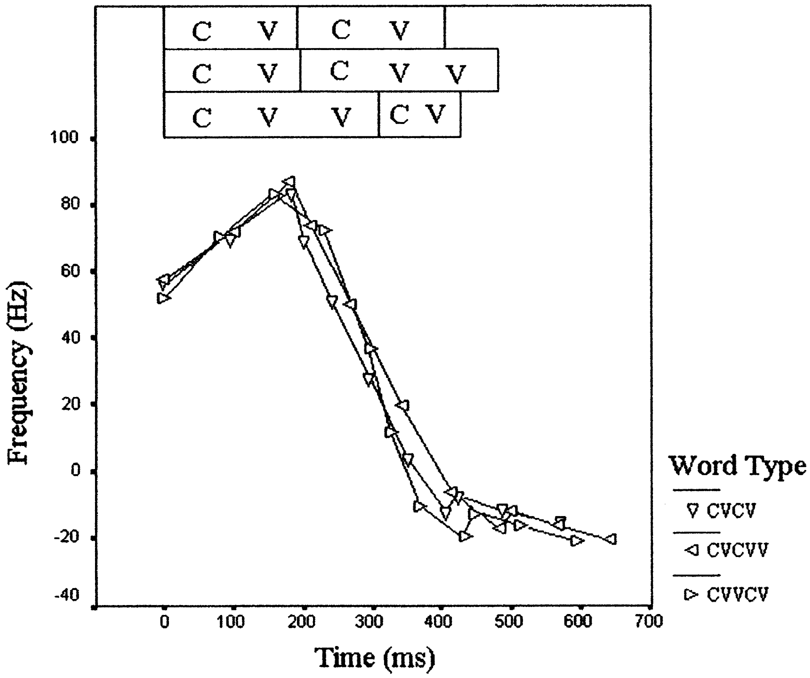

Fig. 4. Mean f_0 values of the Strong Accent versions of the three word structures, at the measurement points employed (for more explanation, see the text). The means (in Hz) are relative to the mean f_0 measured at the end of the preceding *että*. Zero milliseconds on the time scale indicates word onset. For CVCV and CVCVV words the syllable boundary is located between the third and fourth measurement point (approximately at 200 ms from word onset), for CVVCV words between the fifth and sixth measurement point (approximately at 300 ms from word onset). The last three measurement points refer to the initial syllable of the word following the target.

CVVCV (Fig. 2) the apparent turning point is between the third and fourth measurement points, and it too turned out to be similarly reliable: $t1 = 2.65$, $p < 0.05$, df = 9; $t2 = 5.43$, $p < 0.001$, df = 7. In Strong Accent, then, there was a reliable turning point after the end of the first mora in each of the three word structures. In Moderate Accent, the apparent turning point in CVCV (Fig. 1) is located after the fourth measurement point (beginning of the second syllable), and the comparison with the next measurement point proved reliable: $t1 = 4.15$, $p < 0.01$, df = 9; $t2 = 2.65$, $p < 0.05$, df = 10. In CVCVV (Fig. 3) the corresponding comparison produced the same result: $t1 = 4.67$, $p < 0.001$, df = 9; $t2 = 2.91$, $p < 0.05$, df = 7. In CVVCV (Fig. 2), finally, the apparent turning point in Moderate Accent is between the fourth and the fifth measurement point (the latter representing the end of the first syllable), and it too turned out to be reliable: $t1 = 3.44$, $p < 0.01$, df = 9; $t2 = 6.11$, $p < 0.001$, df = 7. In Moderate Accent, then, the turning points were somewhat delayed, in terms of moraic structure, compared to those in Strong Accent, although in real time the turning points appear to have occurred somewhat earlier in Moderate than in Strong Accent. Thus there seems to have been more similarity between the two degrees of accentuation in real time than in terms of moraic structure. A possible interpretation of the situation is that the turning point in Strong Accent occurs earlier than in Moderate Accent, in moraic terms, in order to make sufficient room for the ensuing deep fall (see below).

Since the numerical turning points were real ones also statistically, the next comparison involved the extent of the f_0 rise in the accented versions, from the end of the second syllable of the preceding conjunction *että* to the peak value (which, in effect, amounts to a comparison of the peak values themselves, relative to *että*). In Strong Accent the mean rises were 84.1 Hz for CVCV, 83.4 Hz for CVVCV, and 86.8 Hz for CVCVV, in Moderate Accent 55.2 Hz, 51.2 Hz and 57.0 Hz, respectively. There was a significant effect of Prominence [$F1(1,9) = 14.81$, $p < 0.01$; $F2(1,24) = 171.25$, $p < 0.001$] but no effect of Structure [$F1(1,9) < 0$; $F2(2,24) < 0$] and no interaction [$F1(2,9) < 0$; $F2(2,24) < 0$].

Finally, the mean f_0 values at the end of the target words were compared (again, relative to the preceding *että* conjunction). The effect of Prominence was significant [$F1(1,9) = 9.75$, $p < 0.05$; $F2(1,24) = 96.60$, $p < 0.001$], as was the effect of Structure [$F1(1,9) = 25.95$, $p < 001$; $F2(2,24) = 8.87$, $p < 0.001$], but there was no interaction [$F1(1,9) = 4.61$, ns; $F2(2,24) = 1.39$, ns]. Pairwise comparisons (subjects analysis) indicated that in Strong Accent CVCV had reliably higher f_0 than both CVVCV [$t1 = 3.43$, $p < 0.01$, df = 9] and CVCVV [$t1 = 4.17$, $p < 0.01$, df = 9], and that there was no difference between CVVCV and CVCVV [$t1 = 1.13$, ns]. In items analysis, however, there were no significant differences: the total effect of Structure did barely reach significance [$F2(2,24) = 3.50$, $p = 0.046$] but all Scheffe post hoc tests were nonsignificant. In Moderate Accent CVCV had, in subjects analysis, reliably higher f_0 than both CVVCV [$t1 = 4.94$, $p < 0.001$, df = 9] and CVCVV [$t1 = 3.03$, $p < 0.05$, df = 9], and there was no difference between CVVCV and CVCVV [$t1 = 1.73$, ns]. Items analysis yielded the same result: the effect of Structure was significant [$F2(2,24) = 10.60$, $p < 0.001$], and in Scheffe post hoc tests the CVCV–CVVCV and CVCV–CVCVV differences were significant while the CVVCV–CVCVV difference was not. In Word Stress, finally, CVCV had, in subjects analysis, reliably higher f_0 than CVCVV [$t1 = 2.84$, $p < 0.05$, df = 9], but there was no difference between CVCV and CVVCV [$t1 = 1.92$, ns] nor between CVVCV and CVCVV [$t1 < 0$]. In Items analysis, however, the effect of Structure failed to reach significance [$F2(2,24) = 1.01$, ns]. In Moderate Accent, then, the fall from the peak f_0 value to the end of the target words was the least extensive in CVCV words in which the last segment constitutes the word's second mora, and in Strong Accent and Word Stress this was true numerically.

Turning now to the effect of Prominence, in CVCV Strong Accent involved a reliably lower f_0 than Moderate Accent [$t1 = 3.40$, $p < 0.01$, df = 9; $t2 = 5.88$, $p < 0.001$, df = 10] but the difference between Strong Accent and Word Stress failed to reach significance in the subjects analysis [$t1 = 1.94$, ns; $t2 = 5.48$, $p < 0.001$, df = 10], as did the difference between Moderate Accent and Word Stress [$t1 = 1.98$, ns; $t2 = 5.04$, $p < 0.001$, df = 10]. In CVVCV Strong Accent involved reliably lower f_0 than Moderate Accent [$t1 = 3.62$, $p < 0.01$, df = 9; $t2 = 5.06$, $p < 0.001$, df = 7] and Word Stress [$t1 = 4.33$, $p < 0.01$, df = 9; $t2 = 4.77$, $p < 0.01$, df = 7], but there was no reliable difference between Moderate Accent and Word Stress [$t1 = 1.20$, ns; $t2 = 1.15$, ns]. In CVCVV, too, Strong Accent involved reliably lower f_0 than Moderate Accent [$t1 = 3.76$, $p < 0.01$, df = 9; $t2 = 7.68$, $p < 0.001$, df = 7] and Word Stress [$t1 = 3.28$, $p < 0.01$, df = 9; $t1 = 10.35$, $p < 0.001$, df = 7], and again there was no reliable difference between Moderate Accent and Word Stress [$t1 < 0$; $t2 < 0$]. Thus in each word structure Strong Accent involved a reliably lower f_0 at the end of the word than did at least one of the other degrees of prominence (and in two structures reliably lower f_0 than both of the other degrees). If f_0 at this point were influenced by the f_0 level during the following word, one would rather expect Word Stress to differ from the other two degrees of prominence, as Word Stress (but not the other degrees) was followed by a high rise (signalling contrastive accent). Since this was not the case, we presume that the following tonal context did not influence the f_0 values, and that the values reflect genuine differences among the degrees of prominence. Suffice it to note, without detailed documentation, that at each of the three measurement points along the initial syllable of the word following the target word, Prominence was always fully significant, while Structure had no effect, and there was no interaction. For Strong and Moderate Accent, which were both followed by unaccented words, this means that the difference in the degree of accentuation in fact persisted during the initial syllable of the next word. Thus, there appears to have been no tonal anticipation at the end of the target word (in the sense that the f_0 value at the end of the target word would have been influenced by a value programmed for the next word), and we consequently suspect that a low end point is part of the essence of a Strong Accent as opposed to a Moderate one.

To summarise, a mora-related effect on the extent of the f_0 fall from the peak to the end of the word was observed in Moderate Accent, but otherwise only the degree of accentuation had an effect on the extent of the f_0 rises and falls, irrespective of syllable structure. The mean absolute rise in Strong Accent was 84.8 Hz, from 205.8 Hz (at the end of the preceding *että*) to 290.6 Hz, or 6.0 semitones (ST), that in Moderate Accent 54.5 Hz, from 210.1 Hz to 264.6 Hz (4.0 ST), the mean absolute fall in Strong Accent 102.3 Hz, from 290.6 Hz to 188.3 Hz (7.5 ST), and that in Moderate Accent 55.7 Hz, from 264.6 Hz to 208.9 Hz (4.1 ST). But as shown in Experiment 2 below, the measured rises just given were in fact inflated by an almost 20 Hz step-up unconnected with accentuation.

It can be seen in Figs. 1–3 and Table 2 that in the Word Stress versions of all word structures f_0 was essentially flat during the first syllable, and then fell slightly to the end of the word (statistical comparisons showed no reliable differences between the three word structures, and no reliable f_0 fall or rise, during the first syllable). Except for the slight fall during the second syllable, then, there was thus no tonal movement in Word Stress. We are only guessing—as we have no independent evidence — but the second-syllable fall may be due to the strong contrastive accent on the following word, it may be a lowering whose purpose is to make the following rise stand out more clearly from its immediately preceding tonal context. The initial syllables of the words following the targets were of diverse structures (both mono- and polymoraic), and hence in the strongly accented words following the Word Stress versions of the targets the location of the f_0 peak presumably exhibited the same structure-dependent variation as it did in the target words. Nevertheless, it can be seen in Figs. 1–3 that the f_0 rise following Word Stress target words was considerable in each word structure, indicating that f_0 range was still wide at this late point that was never far from the end of the carrier sentence. That is, flattening of f_0 range cannot explain the flatness of the f_0 curve in Word Stress.[3]

[3]One of the reviewers does not agree with our interpretation of the results concerning the difference between Moderate Accent and Word Stress. Firstly, the reviewer is not convinced that the two degrees of prominence are auditorily distinct, because (a) the auditory judgements were made by someone who was aware of the intended prominence levels, and (b) the tokens were presented (in making the judgements) both in their carrier sentences and in isolation. Thus, the reviewer argues, the evaluator could consciously (or unconsciously) make use of the fact that Word Stress occurred late and Moderate Accent early in the carrier sentence. We answer by reassuring that all three of us hear the difference—a perception that is well supported by the results of our f_0 measurements—even when the word tokens are presented in isolation, without any explicit information concerning the carrier sentence from which a particular

Control analyses were again performed with the five triplets involving only segmentally fully balanced words (e.g., *lama*, *laama*, *lamaa*). Means corresponding to those given in Table 2 were first computed. The average difference between these new means and those in Table 2 was 2.8 Hz (range: 0–11.6 Hz). For the measurement points of the Word Stress versions the average difference was 1.9 Hz (range: 0–4.0 Hz), and for those measurement points of Strong Accent and Moderate Accent that defined the beginning, peak and end values of the rise–fall patterns, the average difference was 1.8 Hz (range: 0.1–6.9 Hz). These last two average differences are both smaller than the overall average difference because the largest differences were observed during the fall portion of the accentual rise–fall, at the measurement points intermediate between the peak and the end that were not included in the statistical tests of the whole material. Next, all differences of means in this smaller material were again tested statistically as was done for the whole material. There were no statistical differences between the two materials concerning Word Stress, and the beginning and peak of Strong Accent and Moderate Accent (including temporal location of the peak). There were a couple of statistical differences concerning the complex pattern of the end values of Strong Accent and Moderate Accent, but these differences did not contradict the conclusions reached above on the basis of the whole material (namely that, in Moderate Accent, the fall from the peak f_0 value to the end of the target words was the least extensive in CVCV words, and that in each word structure Strong Accent involved a reliably lower f_0 at the end of the word than did at least one of the other degrees of prominence).

3. Experiment 2

3.1. Introduction

It was seen in Experiment 1 that there was no reliable f_0 movement during the first syllable in the Word Stress versions of any structural word type. Yet a significant [$t1 = 14.75$, $p < 0.001$, $df = 29$; $t2 = 7.54$, $p < 0.001$, $df = 26$] 20 Hz step-up from the end of the preceding conjunction *että* was observed across the word structures, a step-up that prevents us so far from concluding that word stress was not signalled tonally at all. However, it seemed possible that the step-up represents some kind of boundary tone, and is not a property of word stress as such. Prosodically if not syntactically, the conjunction *että* seems to belong to what precedes it; for example, when pauses occur in the vicinity of *että*, they typically occur *after* the conjunction. It seemed possible that, if a content word unquestionably belonging to the same prosodic unit as the target word had

(*footnote* [3] *continued*)
token has been excised; however, we cannot conclusively exclude the possibility that we unconsciously manage to remember, on the basis of whatever auditory cues that might be available, which carrier sentence each of the 500-odd tokens involved has been excised from. Secondly, the reviewer argues that since the Word Stress versions occurred late and the Moderate Accent versions early in their respective carrier sentences, the possibility cannot be ruled out that the measured differences are due to position in the sentence rather than to level of prominence. We are fully aware that, on logical grounds, our experimental design does not allow the effect of prominence level to be dissociated from the effect of position in sentence, but we nevertheless consider it highly unlikely that the observed differences between Word Stress and Moderate Accent would to any appreciable extent be due to position in sentence; for one thing, it seems incontestable to us that the three-way perceptual opposition between not accented, moderately accented and strongly accented is valid throughout an utterance, from its beginning to the very end.

preceded the target, the step-up would not have occurred on the target word (but possibly on the word preceding it). To test the possibility of such a step-up, Experiment 2 was conducted, in which a content word always preceded the target words. Recall that in Experiment 1 Structure had no significant effect on f_0 at word onset and at f_0 peak. It was therefore decided that one of the three word structures would be sufficient in this test. If the assumption that the step-up was due to a boundary tone is correct, then a similar step-up should now occur on the content word preceding the target, and there should be no step-up between this word and the target in Word Stress.

3.2. Methods

3.2.1. Subjects

Talkers were five volunteers from the same population of logopedics students as the subjects in Experiment 1, but ones who did not participate in that experiment.

3.2.2. Materials and recordings

The same CVCV target words were used as in Experiment 1. For each target word a novel triplet of short sentences was constructed as in Experiment 1, but a content word (an adjective, an adverb, or a genitive attribute) now occurred between the conjunction *että* and the target. In one sentence type (Word Stress) a contrastive accent was again indicated on the verb following the target, the target was mentioned twice, and it was the second occurrence that was measured. An example, involving the target *kato* 'dearth', is "Sanoin että uhkaava kato KAUHISTUTTAA, en sanonut että uhkaava kato KIINNOSTAA" ('I said that the threatening dearth HORRIFIES (one), I didn't say that the threatening dearth INTERESTS (one)'). A Moderate Accent example is "Sanoin että uhkaava kato koskettaisi varsinkin KÖYHIÄ kansalaisia" ('I said that the threatening dearth would concern especially POOR citizens'), and a Strong Accent example is "Sanoin että uhkaava KATO kummastuttaa, en sanonut että uhkaava KAATO kummastuttaa" ('I said that the threatening DEARTH makes one wonder, I didn't say that the threatening FALL makes one wonder'). In a given triplet, the target word was again always followed by the same plosive. There were 11 (targets) x 3 (sentence types or target versions) = 33 different sentences. The sentences were randomised with others designed for another experiment, with the restriction that a given word belonging to this test was not allowed to occur in two consecutive sentences. Five practice sentences preceded the whole randomized list of sentences. The recordings were made as in Experiment 1.

3.2.3. Measurements

One of the authors performed all of the acoustic measurements. F_0 was measured at the end of the conjunction *että* and at the beginning of the following word, at the end of this word and at the beginning, middle and end of the initial syllable of the target word. Another author auditorily evaluated the prominences of the word tokens in their carrier sentences using the same three-way scaling as in Experiment 1: not accented, moderately accented, and strongly accented. In 29 tokens (17.6% of all tokens) the outcome of this evaluation differed from the intended degree of prominence; these differences always involved Word Stress and Moderate Accent.

3.3. Results

Looking first at f_0 movement at the boundary of the target word and the preceding one, the mean rise from the preceding word to the target was 46.5 Hz (from 201.3 to 247.8 Hz) in Strong Accent, 23.0 Hz (from 198.4 to 221.4 Hz) in Moderate Accent, and 3.6 Hz (from 198.6 to 202.2 Hz) in Word Stress. The effect of Prominence on this f_0 rise was significant [$F1(1,4) = 154.65, p < 0.001$; $F2(1,10) = 30.95$, $p < 0.001$]. Pairwise comparisons indicated that Strong Accent differed significantly from both Moderate Accent [$t1 = 4.68$, $p < 0.01$, df = 4; $t2 = 3.19$, $p < 0.01$, df = 10] and from Word Stress [$t1 = 12.44$, $p < 0.001$, df = 4; $t2 = 5.56$, $p < 0.001$, df = 10], and that also Moderate Accent and Word Stress were reliably different [$t1 = 8.13$, $p < 0.001$, df = 4; $t2 = 9.37$, $p < 0.001$, df = 10]. Pairwise comparisons were finally performed, separately for each degree of prominence, on f_0 values obtained at the beginning of the target words and those at the end of the preceding words. The mean rises in f_0 were reliable in each degree of prominence: in Strong Accent [$t1 = 12.50$, $p < 0.001$, df = 4; $t2 = 5.42$, $p < 0.001$, df = 10], in Moderate Accent [$t1 = 9.87$, $p < 0.001$, df = 4; $t2 = 9.87$, $p < 0.001$, df = 10], and in Word Stress [$t1 = 9.06$, $p < 0.001$, df = 4; $t2 = 3.00$, $p < 0.05$, df = 10]. Statistically, then, the results were very similar to those of Experiment 1: There was a reliable f_0 rise at the onset of the target words, even in Word Stress. Numerically, however, the 3.6 Hz (or 0.3 ST) rise in Word Stress is very small compared to the 20.5 Hz rise in Experiment 1. In Experiment 2, too, f_0 was very flat during the initial syllable of the Word Stress versions of the target words (on average 202.2 Hz at the beginning, 202.1 Hz in the middle, and 205.6 Hz at the end), suggesting again that no tonal movement is associated with Word Stress as such. A highly plausible explanation of the 3.6 Hz rise is that it is due to the microprosodic raising effect of voiceless obstruents on the f_0 of the following vowel, even though an attempt was made to avoid this effect in making the measurements. Eight of the target words had an initial voiceless obstruent, and in these words there was a mean rise of 5.1 Hz, whereas in the three targets with an initial (voiced) resonant, there was a mean *fall* of 0.4 Hz.We therefore conclude that the 20 Hz step-up observed in Experiment 1 was due to consequences of the preceding *että* conjunction, and that when content words preceded the target words in Experiment 2, the step-up disappeared.

The same set of 11 CVCV words were used in both Experiments. It is therefore possible that also the size of the microprosodic effect was the same in both. If the step-up due to the conjunction *että* observed in Experiment 1 was inflated by a 3.6 Hz microprosodic effect, then the size of the step-up as such was (20.5−3.6 =) 16.9 Hz. Subtracting this figure from those observed at word onset in Experiment 1 so as to obtain the effects of prominence only, we get a 39.4 Hz (3.0 ST) rise in Strong Accent, a 21.0 Hz (1.6 ST) rise in Moderate Accent, and a 3.6 Hz (0.3 ST) rise in Word Stress. These figures are very similar to the means observed in Experiment 2 (46.5 Hz or 3.6 ST, 23.0 Hz or 1.9 ST, and 3.6 Hz or 0.3 ST, respectively) in the absence of a step-up. It thus seems likely that the mean size of the step-up was indeed somewhat below 20 Hz in Experiment 1.

If the 20 Hz step-up observed in Experiment 1 was indeed a consequence of the preceding *että* conjunction, then a similar step-up should also be observable in Experiment 2, but this time at the beginning of the content word preceding the target. On average, there was a 24.3 Hz step-up from the end of *että* to the onset of the content words. Pairwise comparisons involving all test sentences indicated that this rise was fully significant [$t1 = 8.60$, $p < 0.001$, df = 14; $t2 = 9.05$, $p < 0.001$, df = 32]; notice that the classification into three degrees of prominence did not apply to the content words, only to the target words following them. Thus, Experiment 2 confirmed in two ways that the step-up in Experiment 1 was not a property of word stress: the step-up disappeared from the target Word Stress versions when a content word was added in front of the targets, and the step-up now appeared in the added words.

4. General summary and discussion

The study revealed clearcut differences among the phonetic realizations of three degrees of prominence in Finnish. To our knowledge, no previous studies of the language exist with which the results could be meaningfully compared, as (mere) word stress and accentuation have not been properly distinguished in those—usually much smaller-scale—phonetic investigations that have been reported. Our speakers came from a particular dialect area, and while we suspect that there may be some differences in the precise timing of accentual f_0 movements across different dialects, we presume that otherwise our results are representative of Finnish in general.

It was shown that the main difference between word stress and accentuation is that the former is not realized tonally, while accents are signalled at least tonally, a conclusion that is not at variance with suggestions in recent prosodic models of e.g., English (Cruttenden, 1997) and Swedish (Bruce, 1998). In this respect at least, then, Finnish does not seem to differ from more thoroughly investigated languages. The false assumption that word stress in Finnish is regularly accompanied by a f_0 rise, an assumption due to the interpretation of accents as word stresses, together with the fact that Finnish has fixed word-initial stress, has caused a number of misunderstandings, for example that word onsets in Finnish are always clearly signalled by an f_0 rise; the study by Vroomen, Tuomainen and de Gelder (1998) is a recent example of precisely such a misunderstanding. (Note that the baseline in the present investigation was just word stress, with which accentuation was compared: it will require further investigations to determine what constitutes word stress in Finnish, i.e., what differentiates stressed syllables from unstressed ones.) As expected, Strong Accent differed from Moderate Accent by having more extensive f_0 movements: in addition to the expected higher peak value, Strong Accent also involved a lower end point of the fall than Moderate Accent.

It was observed that the mora unit is important and useful in explaining the course of f_0 movement in Finnish accentuation, the account in terms of morae being much more elegant than one in terms of syllables. In syllabic terms, two distinct, structure-dependent patterns of accentuation were discovered: in one pattern the rise and (at least most of) the fall occurred during the initial syllable, in the other pattern the rise occurred in the initial syllable, the fall in the next one. In moraic terms, the peak of f_0 in strong accentuation invariably occurred at the end of the word's first mora, and in moderate accentuation the peak invariably occurred in the initial part of the second mora, in both cases irrespective of the syllable affiliation of the morae involved. In moraic terms, then, accentuation was uniform across the word structures: rise during the first mora, fall during the second (and eventual following) morae. As for duration, the present results showed moraic effects in strong accentuation: the increase in the duration of V_2 (relative to the mean of the other two degrees of prominence) was considerably longer when V_2 constituted or contained the second mora (on average 41.8 ms or 33%) than when it did not (5.5 ms or 11%).

To be sure, all of the effects described above in moraic terms could also be described without reference to this unit. Thus, e.g., the location of the accentual tonal peak could be characterized in terms of whether the word-initial syllable is light or heavy, or whether its duration is relatively short or long. Nevertheless, we believe that using moraic structure as our frame of reference is justified. First, although descriptive/explanatory elegance is to some extent a matter of taste, we would like to argue that the statement "accentual rise occurs on a word's first mora, and accentual fall on the second one" is more elegant than the statement "in a word with a light initial syllable, accentual rise occurs on the first syllable and accentual fall on the second one, and in a word with heavy initial syllable, both accentual rise and fall occur on the first syllable". (Unfortunately, moraic structure and syllable length (or syllable duration) cannot be dissociated in Finnish, at least as far as mono- and bimoraic syllables are concerned: a bimoraic syllable is always one segment (and also durationally) longer than a monomoraic one.) Second, the distinction between light and heavy syllables is inadequate to capture the facts of Finnish, a quantity language in which syllables can be from one to four morae long (examples are the initial syllables of *su.ti*, *suo.ta*, *suot.ta*, *Suort.ti*); all polymoraic syllables are heavy, and thus reference to syllable weight alone would obscure relevant distinctions (and metrical theories that do not allow syllables with more than two morae are simply unable to describe the basic facts). Third, the motivation for the mora unit is more compelling in stating durational than tonal regularities in the language. In later work (Suomi & Ylitalo, submitted) we have observed e.g., that in (C)VC.CV words (e.g., *kulta*) the first-syllable coda consonant (/l/ in the example) is lengthened much more than the similarly first-syllable coda consonant in (C)VVC.CV words (e.g., *kuulto*), relative to the same consonant in syllable-initial position in (C)V.CV words (e.g., *kulo*). In *kulta* type of words the lengthening was 68%, in *kuulto* type 24% (the lengthening only applied to voiced consonants in such positions, whereas voiceless consonants in the corresponding positions did not exhibit any change in duration or were shortened, a circumstance that is interesting from the perspective of accentuation). In both structures the lengthened consonant constitutes the last segment of the initial long syllable, but what clearly matters is whether or not the consonant constitutes the word's second mora. This is just one example illustrating the usefulness, if not necessity, of the mora unit in accounting for durational regularities in Finnish, a language that is much more clearly mora-counting than e.g., the Germanic languages.

Apart from considerations of moraic structure, the present results on the durational effects of accentuation seem to be similar to those in e.g., English (but for an exception see the next paragraph). Thus Turk and White (1999) found that accentual lengthening in English increases not only the duration of the stressed syllable of the accented word, but the durations of the neighboring syllables within a word as well (with a number of attenuating structural constraints that were not controlled in the present study). This is essentially what was observed in the present study; note that Turk and White only elicited strong (contrastive) accentuation, and in the present study durational effects only appeared in (contrastive) Strong Accent.

In contrast to the study by Turk and White (1999), in which no lengthening effects were observed to the left of an accented syllable across a word boundary (in Scottish English), it was observed above (in Experiment 1) that the duration of the preceding *että* increased incrementally as a function of increasing prominence on the target words (irrespective of word structure). At the moment at least, we can do little more than note some differences between the two studies. Apart from the fact that different languages are involved, another difference is that *että* is a function word (and a conjunction at that), while in Turk and White's study (in their Experiment 2) the words preceding the accented targets were content words. The results of our experiments suggest that the conjunction, before each degree of prominence of the target words, was followed by a prosodic boundary of some kind, which in turn implies that the observed effect may be due to preboundary lengthening, but we cannot explain why the extent of such lengthening should vary as a function of the degree of prominence of the following word. Clearly, the behavior of conjunctions (and function word in general) with respect to accentual lengthening deserves further investigations within and across languages.

Assuming that the postpeak fall is as essential a part of accentuation as the preceding rise to the peak, the distinction drawn by Ladd (1996, p. 55) between *association* and *alignment* comes in handy: normal accents in Finnish are unquestionably associated with the initial, stressed syllables, but their *alignment* with the segmental material follows the moraic pattern just summarised ("normal" accentuation excludes instances in which a contrastive accent falls on a noninitial syllable). It is this unusual discrepancy between (abstract and structural) association and (concrete and realizational) alignment that makes foreign listeners hear stress/accent on the second syllable in words with monomoraic initial syllable. In Swedish, the distinction between Accent I and Accent II also involves a timing difference in the alignment of the f_0 fall with the segmental material, but the Swedish difference of alignment is not conditioned phonologically (Bruce, 1977).

It was noted in the Introduction that, according to Hayes (1995), Finnish is typologically rare in its specific combination of stress (or rhythmic) system and quantity system: Finnish is exceptional among the syllable trochee languages in having a quantity distinction, and in that syllable weight nevertheless plays no part in the assignment of (primary) stress. Extending the term stress, as used by Hayes, to encompass accentuation as well, it can be stated that Finnish indeed seems to be exceptional in the sense intended by Hayes as far as the association of accentuation is concerned, but the exceptionality is attenuated when the phonetic alignment (relative timing of the f_0 contour with the segmental string) is considered. Thus when the first syllable is light (or short), the phonetic realization of accent is extended to the second syllable (which realizes the fall portion of the accent); in a sense, then, a light initial syllable alone is unable to constitute the phonetic domain of an accent, and thus syllable weight does play a part, namely in the alignment of stress/accent with the segmental material.

Although reference has been repeatedly made above to rises and falls of f_0, we are inclined to view the observed (phonetic) f_0 movements as transitions or interpolations between (phonological) level tonal targets, in line with the framework of the bitonal (or autosegmental-metrical) approach to intonational phonology, originating in the work of e.g., Bruce (1977) and Pierrehumbert (1980). A strong empirical reason for this inclination, for us, is the constancy that was observed in the extents of rises and falls in the two degrees of accentuation, irrespective of word structure. In both degrees of accentuation, the beginning and end values of the f_0 rise were statistically identical across word structures, and similarly for the beginning and end values of the f_0 fall (except that, in Moderate Accent, the fall from the peak f_0 value to the end of the target words was the least extensive in CVCV words in which the last segment constitutes the word's second mora). In the present material, however, foot and word cannot be dissociated as all target words consisted of one foot. Future investigations should vary word length and foot structure independently.

Acknowledgements

We wish to thank Alice Turk, Ilse Lehiste and Mattias Heldner for very useful comments on earlier versions of this paper. After such a thorough and expert review, any errors and misinterpretations that remain are definitely our own. We thank Pentti Körkkö for technical assistance, and the third author gratefully acknowledges the grant from the Faculty of Humanities at Oulu University.

Appendix 1

List of the target words grouped according to structure.

CVCV	CVCVV	CVVCV
kato		kaato
kesä	kesää	
kina	kinaa	Kiina
kota	kotaa	
lama	lamaa	laama
latu		laatu
lima	limaa	liima
pata	pataa	
sika	sikaa	siika
sima	simaa	siima
tuli		tuuli

References

Berinstein, A. (1979). A cross-linguistic study on the perception and production of stress. *UCLA working papers in phonetics 47*, University of California, Los Angeles.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.

Bruce, G. (1998). *Allmän och svensk prosodi*. Praktisk lingvistik 16. Institutionen för lingvistik, Lunds universitet.

Collinder, B. (1937). Über Quantität und Intensität. *Neuphilologische Mitteilungen* 1937, 97–120.

Cruttenden, A. (1997). *Intonation* (2nd edn). Cambridge: Cambridge University Press.

Hayes, B. (1995). *Metrical stress theory*. Chicago: The University of Chicago Press.

Huss, V. (1978). English word stress in the post-nuclear position. *Phonetica*, *35*, 86–105.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehtonen, J. (1970). *Aspects of quantity in standard Finnish*. Jyväskylä: Jyväskylä University Press.

Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation*. Ph.D. thesis, MIT.

Sadeniemi, M. (1949). *Metriikkamme perusteet*. *Suomen kirjallisuuden seuran toimituksia*, vol. 236. SKS: Helsinki.

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

Suomi, K. (1980). *Voicing in English and Finnish stops*. Publications of the Department of Finnish and General Linguistics of the University of Turku 10.

Suomi, K., & Ylitalo, R. (submitted). On durational correlates of word stress in Finnish.
Turk, A., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, *27*, 171–206. doi: 10.1006/jpho.1999.0093.
Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, *38*, 133–149.