Simple Perception Experiments on Estonian Word Prosody: Foot Structure vs. Segmental Quantity*

Arvo Eek and Einar Meister

Abstract

The following paper comprises three main parts. In the initial part we present a concise review of the data concerning Estonian word prosody. The data found in the literature raise new questions. In the second part some of these questions are tested by perception experiments. We succeeded in showing the existence of only two (short vs. long), not three phonological quantity degrees on segmental and syllabic levels. All three accents (traditionally called syllabic quantity degrees) can be identified only when information about the second syllable vowel has been delivered to listeners. The smallest prosodic unit in Estonian does not coincide with the syllable. Disyllabic accent feet are minimal distinctive prosodic units in Estonian. The main region of the stress foot, essential for the perception of accent types, consists of the stressed syllable rhyme (minimally its final part, i.e. the final half of the long nucleus or the coda peak) and of the nucleus of the following unstressed syllable. The concentration of the perceptually relevant information around the syllable boundary creates a perceptual impression that the long heavy accent (A3) has always a certain kind of peakedness whereas the long light accent (A2) has not. Besides segmental durations, listeners need additional information for the identification of accents embedded in a wider context than is delivered by the stressed syllable alone. All A3 feet are either vowel-peaked or consonant-peaked, other possibilities are lacking. In an unbalanced A3 disyllabic foot the perceptual effect of peakedness is strengthened against the background of the weakened, shortened, and qualitatively reduced second syllable vowel. The balance of A2 becomes conspicuous against the background of the lengthened and unreduced second syllable vowel. Accents, as the minimal distinctive prosodic units, are closely connected with the stress pattern of a word. The domain of accents is a stress foot, irrespective of whether the foot is primarily or secondarily stressed. In the stress foot there operate two temporal regulation mechanisms. The higher-level temporal regulation, determined by the stress for the whole stress foot (up to trisyllabic feet), is expressed in an isochronic trend according to which addition of phonemes to the stress foot decreases the durations of the constituent segments in the foot, but does not affect accents. The lower-level temporal regulation mechanism, subordinated to the whole stress foot, characterizes isochrony of the accent foot. A stress-conditioned trend to isochrony influences absolute durations of the intrasyllabic constituents, whereas an accent-conditioned temporal regulation (e.g. inversely proportional duration relations) expresses intersyllabic accent-characteristic temporal reorganization of the constituents in isochronous accent feet. In the final part of the paper we present the main principles of a conceptual framework for the description of Estonian word prosody.

1. Introduction: A Review with Questions

Nowadays, two hierarchical conceptions of word prosody prevail in the Estonian phonetic-phonological literature (Eek 1994a).

* The support from Estonian Science Foundation has made the present research possible.

Ilse Lehiste and Jaan Ross (editors), Estonian Prosody: Papers from a Symposium, pp. 71–99.

The more traditional treatment is based on three distinctive degrees of segmental duration, i.e. all vowels and consonants, with some exceptions, occur in short (Q1), long (Q2) and overlong (Q3) distinctive segmental duration degrees (I, II, III *välde*, resp.; e.g. Ariste 1939, 1953; Lehiste 1960). In transcription the segmental degrees are differently expressed depending upon phonological interpretations of those degrees:

Q1 – V, C

Q2 – $\bar{V}$, $\bar{C}$; VV, CC; V:, C:

Q3 – $\hat{V}$, $\hat{C}$; VVV, CCC; V::, C::

The inversely proportional relations between the sound durations in adjacent stressed/unstressed syllables (Posti 1950; Lehiste 1960) have provided a phonetic rationale to examine the duration degrees also on higher, syllabic level: stressed syllables have three distinctive degrees of syllabic quantity (denoted also by Q1, Q2 and Q3; Lehiste 1960). The hierarchical treatment is inevitable also due to the fact that the so-called Q2/Q3 diphthongs and consonant clusters act similarly to the Q2/Q3 monophthongs and geminate consonants in quantity phonology. In monosyllabic words the Q2/Q3 contrast is neutralized. The archidegree in the neutralization position is conventionally represented by Q3 (Hint 1977).

Binary treatments (e.g. Trubetzkoy 1939; Posti 1950; Ravila 1962; Hallap 1963; Harms 1962, 1978; Vihman 1974), differing in details and interpretations, deny the three-way quantity contrast on segmental level. The three-way quantity contrast is hierarchically divided into two binary contrasts. (1) All vowels and consonants are either short or long. Also in this case the transcription of the short/long segmental contrast depends on the interpretations of phonemic units:

short – V, C

long – $\bar{V}$, $\bar{C}$; VV, CC

(2) The feet, beginning with a long stressed syllable (i.e., in traditional terms, Q2 and Q3 syllables), are contrasted as bearing two distinctive accents — the light and the heavy accent. The feet, beginning with a short (open) stressed syllable, carry the light accent (Tauli 1973; Viitso 1981; Eek 1986). Here and henceforth we designate the accents as follows:

A1 – short light accent, i.e. the foot beginning with a short stressed syllable;
A2 – long light accent, i.e. the foot beginning with a long stressed syllable;
A3 – heavy accent, i.e. the foot beginning with a long stressed syllable.

We proceed from the fact that the degrees of quantity are closely connected with the stress pattern of a word. The domain of the accents (quantity degrees) is a foot. Every foot has a certain accent, irrespective of whether the foot is organized by primary or secondary stress (but cf. Hint 1973; cf. also Lehiste 1965). The difference between A1 and A2 is based on duration (Eek 1980a, b). Such an approach assumes that there can also be other features besides segmental duration that may be taken into consideration especially in the identification of A2 and A3.

Which of these two conceptions reflects more adequately the essence of the Estonian word prosody? Although the data found in literature seem to support the accentual character of quantity degrees, they also raise new problems. It is even not quite clear what the phonetic cues are that constitute the basis for the perception of the accents.

We test by perception experiments the hypothetical answers to the questions raised in the following review.

(1) How many distinctive segmental duration degrees are there in Estonian? Three or two? Even today, this is the main question.

Do we have convincing affirmation of three distinctive segmental duration degrees? Yes and no. Only intervocalic consonants (especially after short vowels) support the traditional ternary system of segmental quantity degrees. But the durational difference between Q2 and Q3 vowels is smaller and inconsistent. These vowels are frequently durationally overlapped, mainly due to especially large standard deviations of Q3 vowels (Eek 1975: Tables 3–5; 1986: Fig. 3). At the same time the durational difference between short and long (traditional Q2 and Q3 taken together) vowels and consonants is consistent and sufficiently persuasive.

Grouping duration data by phonemic and morphological composition of words (e.g. Eek 1986: 14; 36–37), we can register seven phonetic duration degrees for vowels and at least four degrees for intervocalic consonants, as shown in Tables 1 and 2. Let us mention here that the Q2 vowel in A2 foot *saate* 'you get', 2.pl.pres., and the Q3 vowel in the consonant-peaked A3 foot *loota* 'to hope', da-infinitive, are both in diminished long phonetic degree, i.e. they are shorter than the Q2 vowel in the A2 foot *saade* 'sending', nom.sg. The Q3 consonant in the vowel-peaked A3 foot *loota* 'story', abess.sg., and Q2 consonants in Q2 feet *kata* 'cover', 2.sg.imperat., and *saate* 'you get', 2.pl.pres., are all in the long phonetic degree, whereas the Q3 consonant in the consonant-peaked A3 foot *loota* 'to hope', da-infinitive, is usually shorter (in one-and-a-half-long phonetic

Table 1. Average segmental durations in milliseconds for vowels (male speaker). A bold letter marks a sound in the corresponding phonetic duration degree. Examples are in Estonian orthography.

extra short vaeglühike	short lühike	halflong poolpikk	diminished long vaegpikk	long pikk	one-and-a-half-long poolteist-pikk	overlong ülipikk
60	100	140	175	200	240	300
saade[1], A3	sade[2], A1 saade[3], A2 valede[4], A1 valet[5], A1 valeta[6], A1	sade[2], A1 valed[7], A1 valede[4], A1	saate[8], A2 loota[9], A3	saade[3], A2	saade[1], A3 loota[10], A3	maagi[11], A3

saade[1] 'haycocks', part.pl.; *sade*[2] 'sediment', nom.sg.; *saade*[3] 'sending, broadcast', nom.sg.; *valede*[4] 'lies', gen.pl.; *valet*[5] 'lie', part.sg.; *valeta*[6] 'lie', 2.sg.imperat. or 'lie', abess.sg.; *valed*[7] 'lies', nom.pl.; *saate*[8] 'you get', 2.pl.pres.; *loota*[9] 'to hope', da-infinitive; *loota*[10] 'story', abess.sg.; *maagi*[11] 'even the land', nom.sg. + particle -gi.

degree) than the Q3 consonant after a short vowel in the A3 foot *pattu* 'sin', part.sg. Phonologically long consonants between the second and third unstressed syllables in a trisyllabic foot (generally: everywhere after the nucleus of unstressed syllables, irrespective of the type of foot accent) are in the ambiguously long phonetic degree. This phenomenon is one of probable features signalling unstressedness of the syllable (cf. e.g. *'valeta* 'lie', 2.sg.imperat. or 'lie', abess.sg., A1 foot, and idiosyncratic variants *'kava,lamate* 'slyer', gen.pl., both feet in A1, or *'kavala,mate* 'slyer', gen.pl., A1 foot + A2 foot).

Table 2. Average segmental durations in milliseconds for consonants (male speaker). A bold letter marks a sound in the corresponding phonetic duration degree. Examples are in Estonian orthography.

short	long	one-and-a-half-long (ambiguously long acc. to Lehiste)	overlong
60	120	140–160	180
pada[1], A1	kata[5], A2	loota[8], A3	pattu[10], A3
saade[2], A2	saate[6], A2	valeta[9], A1	
saade[3], A3	loota[7], A3		
valede[4], A1			

pada[1] 'kettle', nom.sg.; *saade*[2] 'sending, broadcast', nom.sg.; *saade*[3] 'haycocks', part.pl.; *valede*[4] 'lies', gen.pl.; *kata*[5] 'cover', 2.sg.imperat.; *saate*[6] 'you get', 2.pl.pres.; *loota*[7] 'story', abess.sg.; *loota*[8] 'to hope', da-infinitive; *valeta*[9] 'lie', 2.sg.imperat. or 'lie', abess.sg.; *pattu*[10] 'sin', part.sg. or illat.sg.

Obviously we should discard the opinion that there are four distinctive quantity degrees (accents) in Estonian (see discussions : Remmel 1975; Eek 1977, 1983; Viitso 1981; Hint 1983). The phonetic phenomena set forth as Q4 can be interpreted as variants of the heavy accent predetermined by the morphological structure of the word and by its position in the sentence (see data: Eek 1983: 553–558).

The problem is how the native listener perceives the phonetic duration degrees as suprasegmental distinctive units. If we still wish to hold on to the opinion that there are three distinctive segmental duration degrees in Estonian, then it should be possible to test these segmental degrees by perception experiments like any other phonemic unit which can be identified by a speech segment consisting of no more than two adjacent phonemes. Presumably listeners are not able to identify, on the basis of the sequence consisting of adjacent V and C alone, the distinctive duration degree of the corresponding vowel or consonant. It does not mean that they cannot determine in the pairs of VC+VC sequences which vowel or consonant of the pair is longer or shorter. We do not preclude this kind of discrimination ability; we doubt the ability to identify the distinctive Q's without knowing information from the next syllable. Listeners can, however, identify the binary short/long segmental contrast, comparing durations of the vowel and the following consonant. It is clearly seen in monosyllabic words (cf., e.g. *kuub* 'coat', nom.sg. – *kupp* 'blister', nom.sg. – *kuup* 'cube', nom.sg.). But what about disyllabic words? We will test these problems below.

(2) All researchers have mentioned that the duration ratio of syllables (or V1:V2) is a more consistent feature characterizing word accents than the duration of individual segments. Krull's data from conversational speech (Engstrand, Krull 1994) manifest the V1 overlap in A2 and A3 words for one speaker in particular, but the V1:V2 ratios "separate quantity degrees better than the durations of V1 or V2 taken individually". Duration ratios from different authors are presented in Table 3.

Table 3. Duration ratios in Estonian disyllabic sequences.

	A1 (Q1)	A2 (Q2)	A3 (Q3)
Lehiste 1960	0.7	1.5	2.0
Liiv 1961	0.7	1.6	2.6
Eek 1974	0.7	2.0	3.9
Krull 1991,1992	0.5-0.7	1.2-2.1	2.2-2.9

The results of perception experiments have confirmed that for the identification of accents the duration ratio is a more essential cue than absolute durations of segments (e.g. Lehiste 1975, 1988; Eek 1980a, b). Lehiste (1988: 82) notes "that duration of the first syllable is not sufficient for the perception of a particular accent: if items that have different first syllable durations, but the same ratio, are perceived as being the same, the listeners are responding to the whole disyllabic sequence rather than to the duration of the first syllable alone".

But now we have difficulties using these data for explaining the perception judgements. Namely, the experiments on the perception of duration ratios by Fox and Lehiste (1987, 1989) have shown that listeners recognized only two contrastive patterns: 1:2 and 2:3 (a typical A1 ratio) vs. 3:2 (a typical A2 ratio) and 2:1 (A3), i.e. they discriminated only the cases where the second syllable is longer than the first syllable from the cases where the second syllable is shorter than the first syllable. On the basis of ratios listeners can discriminate A1 from A2 and A3, but not the sequences of syllables typical of A2 and A3. What could be the reason of this contradiction: on the one hand it is accepted that the duration ratio of the syllables is an important cue for the identification of accents, but on the other hand we know that listeners do not discriminate a typical A2 duration ratio from that of A3? Which segments' duration ratios, differing diametrically and hence being perceivable, could listeners actually use instead of the non-distinguishable syllabic duration ratios in the identification of A2 and A3? Perhaps the mentioned contradiction could be overcome if we take into account the durations of the final part of the stressed syllable and of the vowel of the unstressed syllable, i.e. the durations of the segments that seem to be most important in the identificaton of accents (Eek 1986: 32–34). Such a duration ratio treatment would be reasonable if one could prove that the difference between A2 and A3 is also perceived on the basis of the final part of the stressed syllable and the vowel of the unstressed syllable, so that listeners would not necessarily need the whole phonemic information contained in the disyllabic sequence.

(3) The problems concerning duration ratios, treated in (2), are related to isochrony. In the Estonian language a strong temporal compression can be noticed: mono- and

disyllabic feet, irrespective of the accent type and segmental duration, differ durationally less than intra-feet segments. Although generally a disyllabic A1 word is the shortest, there are enough examples where the total duration of A1 does not differ from the total durations of the disyllabic units of A2 and A3. The duration of the A2 and A3 words (however, read in laboratory conditions in frame sentences) did not differ essentially. Thus the tendency of overlapping of total durations of A2 and A3 disyllabic feet is not limited only to reading poems written in the trochaeic metre (cf. Lehiste 1995).

There exists a significant temporal relationship among all segments that constitute a foot (Lehiste 1972; Eek 1974). The word has been regarded as an isochronous unit of a temporal programme (Shockey, Gregorski, Lehiste 1971). Intra-word feet are somewhat shorter than the same feet constituting separate words. However, we have reason to assume that in longer words the foot is the main unit of isochrony. An indirect proof for this is the fact that the universal phenomenon of the duration of the stressed syllable nucleus being inversely proportional to the word length, is restricted in longer words by the foot structure of the word (Eek, Remmel 1975). In words like *padu* 'a low wet place', nom.sg., *padus* 'in the low wet place', iness.sg., *padust* 'from the low wet place', elat.sg., *padusta* 'change into the low wet place', 2.sg.imperat., *padustama* 'to change into the low wet place', ma-infinitive, *padustamatu* 'the piece of land that is impossible to change into a low wet place', nom.sg., *padustamatut* 'the piece of land that is impossible to change into a low wet place', part.sg., *padustamatute* 'pieces of land that are impossible to change into a low wet place', gen.pl., *padustamatutes* 'in the pieces of land that are impossible to change into a low wet place' iness.pl., *padustamatutesse* 'into the pieces of land that are impossible to change into a low wet place', illat.pl., *padustamatutessegi* 'even into the pieces of land that are impossible to change into a low wet place', illat.pl. + particle *-gi*, the duration of the primary-stressed syllable shortens with increasing the word length up to the secondary-stressed syllable *-ta-* in the word *padustama*. A further increase in the word length is not accompanied by the regular shortening of the primary-stressed syllable. Meanwhile, the absolute duration of the secondary-stressed syllable *-mat-* (*padustamatu*) decreases regularly with the lengthening of the word up to the next foot which begins with the syllable *-tes-* (*padustamatutesse*) carrying a new secondary stress. In agglutinative languages with very long words the shortening effect functions only within the limits of the foot, creating an impression as if a long word consists of 2–3 separate words. Although Krull (1991, 1992) has measured the effect of adding syllables only on the duration of the vowels of the first foot, it can be seen from her results as well that the vowel of the first syllable shortens strongly until the word has acquired four syllables and has been restructured into a sequence of two metric feet. From that on the shortening of the vowel of the first syllable is neither remarkable nor regular.

The foot is an integral stress unit. This may be the very reason why all the isochronic phenomena within a foot are not important to the same degree in defining the accent type of a foot. As we could see from the data presented above, the duration of a stressed syllable vowel shortens when consonants are added to the end of an unstressed second syllable and it shortens even more when a whole unstressed third syllable is added to the foot. The whole third syllable is inert in defining accents. It seems as if we were dealing, on the one hand, with stress isochrony which comprises the whole stress

foot, and on the other hand, with isochrony characterizing the accents. Wiik (1985:133) has noted about the latter: "...die erste Silbe und der Vokal der zweiten Silbe (also nicht die ganze Silbe) bilden eine isochronische Einheit". This genuine Lappo-Finnic foot isochrony works even today. According to Wiik (1995), "Foot was originally defined as the sequence from the beginning of the first vowel to the end of the second syllable vowel". Above we have suggested that in the identification of A2 and A3 it is probably sufficient for a listener to compare the end of the stressed syllable with the vowel of the unstressed syllable, i.e. to compare the segments of the smallest isochronic unit in the frame of the largest unit, the stress foot. But it would be really possible if one presumes that short and long segments have been determined earlier during the first stage of perception.

(4) In perception experiments where modified natural base words were used as stimuli, an asymmetry in judgements was registered: it was more difficult to get a natural-sounding A2 word from an A3 base word than vice versa (Eek 1980a, b). Lehiste's results also reveal a certain singularity of the heavy accent (Lehiste 1988:85): "It is thus clear that F0 contour is crucial in the identification of Q3 even under optimal duration conditions. But still — even with the addition of F0, the highest identification scores for Q3 reached only 51% ..." Both researchers have arrived at the same opinion that apparently something else is needed — something that was not present in the synthesized signal in Lehiste's experiments and that was not modified in Eek's experiments with natural speech.

Eek (1986) has mentioned that the different distribution of articulation energy in the quantal unit of energy (i.e. in the foot) may also play a role in the identification of A2 and A3. It was presumed that the articulation energy of A2 is balanced between two syllables. In an A3 foot the energy is unbalanced. It is centered at the end of the stressed syllable rhyme (in vowel-peaked feet *saada* 'to get', da-infinitive, *loota* 'story', abess.sg., *võita* 'without butter', abess.sg. — at the end of the nucleus, and in consonant-peaked feet *patta* 'kettle', illat.sg., *loota* 'to hope', da-infinitive, *võita* 'to win', da-infinitive — at the coda peak; Eek 1986:32). This treatment is similar to the contrast *controlled* vs. *ballistic* presented by Harms (1978) and is quite close to what Ravila (1962:491) has said on Lappish word forms: "... it can be said that their total quantity is largely the same, and equal amounts of articulation energy have been spent on them. The only difference is that quantity and intensity are differently divided on the different phonemes".

Yet this aspect has been studied acoustically and perceptionally least of all. The data presented in Remmel, Eek (1971) confirm the different distribution of the spectral energy of geminate laterals in A2 and A3 words. On the basis of these differences we have set up a binary tense/lax contrast for A3 and A2 (see also Vihman 1974). It should be added: in A2 words, the average intensity of the syllable-final first component of geminate sonorants is 4.8 dB relative to an arbitrary standard, and 3.5 dB of the syllable-initial second component; the corresponding values for A3 words are 4.0 and 0.5 dB (Eek 1994b).

Balanced vs. unbalanced articulation energy is obviously expressed by the total effect of several acoustic features. Probably those features have a different weight depending on the foot type and on its position in the sentence. Our data characterize words from laboratory speech in the sentence stress position.

The stressedness of the short first syllable of A1 is balanced by the half-length of V2 (the phonologically short V2 can be phonetically long). V2 has no quality reduction and its intensity equals to or is even bigger than that of V1. The F0 peak is achieved at the end of V1; during V2 F0 is steadily falling, but the fall starts on a higher frequency than in other accents. In an open first syllable of A2 (e.g. *saada* 'send', 2.sg.imperat.) the F0 peak is achieved in the middle or in the second half of V1 and stays level to the end of the vowel (rising or rising-level contour). During V2 F0 falls, starting the fall on a lower frequency than in A1. The intensity acquires its peak value at the end of V1 and maintains almost the same level during the short or even halflong V2. In a closed first syllable of A2 (e.g. *samma* 'sturgeon', gen.sg.) the F0 peak is achieved in the last quarter of V1 and F0 maintains almost the same height throughout the final part of the vowel and even at the beginning of the first component of the geminate. The impression of the balanced first syllable end and the second syllable is, perhaps, caused by the fact that both components of the geminate are almost of the same intensity, and that the intensity of the qualitatively unreduced short or halflong V2 is not essentially weaker than that of V1. In a vowel-peaked first syllable of A3 (e.g. *saada* 'to get', da-infinitive) the F0 peak is reached during the first third or the first quarter of V1, where a steady fall of F0 starts (falling contour). The fall continues during V2, beginning from a still lower frequency than in A2. The V2 intensity is 5 or more dB weaker than in V1, and an extra short V2 is qualitatively strongly reduced. In a consonant-peaked syllable of A3 (e.g. *samma* 'the same', illat.sg.) the F0 peak is reached in the middle of V1; the F0 fall begins before the end of V1 and continues up to the end of the word. The first component of the geminate that ends the first syllable is stronger than the second component that starts the second syllable. The intensity of an extra short and qualitatively reduced V2 can be 5–10 dB weaker than that of V1. Thus the difference in articulation energy between the end of the first syllable and the second syllable in A3 words is noticeably expressed by acoustic features.

Below we will study the role of fundamental frequency, intensity and duration in the identification of A2 and A3. The main purpose of these perception experiments was to examine under which conditions it could be possible to change an A3 word to a natural-sounding A2 word and vice versa.

2. Methods and Material

One male speaker (from the western dialect area) read 19 disyllabic words in frame sentences "*Ütle ... taas*" 'Say ... again' or "*Ütle ... veel*" 'Say ... again' in the soundproof studio of the Laboratory of Phonetics and Speech Technology in Tallinn. The speech samples were digitized at 16 kHz. Spectrograms, F0 contours and energy curves were computed and their modifications for the perception experiments were performed using the Kay Elemetrics CSL 4300 system. Measurement results of the base words are presented in Tables 4 and 5. In Table 4 the duration of the second component (M2) of a long vowel VV was obtained by subtracting the short first-syllable vowel of the A1 word, pronounced in the same paradigm, from the whole duration of a long vowel. In Table 5 the duration of the first component (coda peak) of geminate consonants was obtained by subtracting the intervocalic short consonant of the A1 word, pronounced in the same paradigm, from the whole duration of a geminate.

3. Notes on the Phonological Transcription

In the present transcription all long vowels and consonants are designated by sequences of two short phonemes. There is no elegant and satisfactory solution (see also Viitso 1981, 1986). Treating long consonants as geminates (although it is not possible to ascertain rearticulation within a geminate: Eek 1976) will probably raise less opposition than the parallel solution for vowels. This opinion is based on the fact that the second component of geminate consonants always belongs to the onset of the next syllable and is therefore inert in accent phonology (just as the second consonant of the two-consonant cluster). Such a transcription mode, based on speech production data, is also supported by Standard Estonian palatalization. However, one could say that the sequence of two short consonants does not constitute even a cluster of two identical palatalized consonant phonemes, because in geminates (and consonant clusters) only the final part of the stressed syllable is palatalized, while the part belonging to the onset of the next syllable is depalatalized according to the articulation place of the following vowel, e.g. /pat:te/ 'sins', part.pl, and /pat':te/ 'stalemates', part.pl. (Eek 1973). It could be felt strange to transcribe the word-final long consonant as a geminate. But also here, justification for the general approach could be found. As it was said, the final part of a long intervocalic (also palatalized) consonant, i.e. the second component of a geminate, manifests the articulatory transition to the following sound, while the second component of the word-final geminate, unreleased in very rare cases, represents the movement to neutral position or to the next word (e.g. in the words /pat:t/ 'sin', nom.sg., and /pat':t/ 'stalemate', nom.sg., the second component is in both cases perceived alike: *pat:t^{o}h* and *pat':t^{o}h*).

Although the variant VV is in accordance with the treatment of diphthongs in accent phonology, this solution is considered inadvisable due to its psychological unreality: geminate vowels (unlike geminate consonants) are said not to be perceivable. The counterargument has not been chosen successfully. There is some evidence that the speaker has an awareness of long vowels consisting of two components. It should be mentioned here that long vowels are diphthongized in some dialects (e.g. *töö* 'work', nom.sg. > *tüö*, *tee* 'road', nom.sg. > *tie*, etc.). According to Lehiste (1985) the *bi* word game confirms monophonemicity of a long vowel and biphonemicity of a diphthong. Yet the same game could support also the biphonemicity of a long vowel because the player always places the *bi*-syllable after the first component of the long vowel.

In accent phonology the syllable and foot treatment cannot be avoided. Replacing $\bar{V}$ and $\bar{C}$ with VV and CC, in spite of the "inadvisabilities" mentioned above, makes it possible to present syllable and foot structures more adequately.

In the transcription system used in this article, a colon does not denote a Q3 phoneme but indicates that the whole foot bears the heavy accent. The position of the colon distinguishes the vowel-peaked foot types from consonant-peaked types (cf. also Hallap 1963; Viitso 1963; Hint 1966). In a vowel-peaked A3 foot, the colon is located after the second component of VV (or V_1V_2). In a consonant-peaked A3 foot, the colon is located after the peak component of a geminate or a consonant cluster, i.e. after the energy center of the unbalanced heavy accent (generally, after the last mora in the

stressed syllable of the metric foot of the heavy accent). The feet without colon bear the light accent.

Although the location of the A3 peak is predicted by rules (about the rules: Eek 1986: 35–36), we use the system with a redundant colon in order to visualize the studied structures. The system where A1 and A2 are marked with acute /´/ and A3 with gravis /`/ before the nucleus of a stressed syllable, is free from the mentioned redundancy. But the weakness of this system is that a juncture should be marked more frequently than with the colon system, cf. e.g.:

/sata/, /saata/, /saa:ta/ – /s´ata/, /s´aata/, /s`aata/
sada 'hundred', nom.sg., *saada* 'send', 2.sg.imperat., *saada* 'to get', da-infinitive
/lauta/, /lau:ta/, /autto/, /laut:ta/ – /l´auta/, /l`auta/, /´autto/, /l`autta/
lauda 'cowhouse', gen.sg., *lauda* 'table', part.sg., *auto* 'car', nom.sg., *lauta* 'cowhouse', part.sg.
/tappa/, /tap:pa/ – /t´appa/, /t`appa/
tapa 'kill', 2.sg.imperat., *tappa* 'to kill', da-infinitive
/karta/, /kar:ta/, /korvi/, /kor:vi/ – /k´arta/, /k`arta/, /k´orvi/, /k`orvi/
karda 'fear', 2.sg.imperat., *karda* 'metal foil', part.sg., *korvi* 'basket', gen.sg., *korvi* 'basket', part.sg.
/nartta/, /kart:ta/, /harffi/, /harf:fi/ – /n´artta/, /k`artta/, /h´arffi/, h`arffi/
narta 'dogsledge', nom.sg., *karta* 'to fear', da-infinitive, *harfi* 'harp', gen.sg., *harfi* 'harp', part.sg.
/kõrtsi/, /kõrt:si/, /vint:skle/ – /k´õrtsi/, /k`õrtsi/, /v`intskle/
kõrtsi 'tavern', gen.sg., *kõrtsi* 'tavern', part.sg., *vintskle* 'writhe', 2.sg.imperat.
/lootte/, /loot:ta/ – /l´ootte/, /l`ootta/
loote 'you create', 2.pl.pres., *loota* 'to hope', da-infinitive
/loo:tta/, /loo:tki/, /maa:kke/ – /l`oo,tta/, /l`oot,ki/, /m`aa,kke/
loota 'story', abess.sg., *loodki* 'even the stories', nom.pl. + particle *-ki*, *maake* 'small country', nom.sg.
/kaartte/, /kaart:te/ – /k´aartte/, /k`aartte/
kaarte 'windrows', gen.pl., *kaarte* 'cards', part.pl.
/pois:se/, /poi:sse/, /kuus:ki/, /kuu:ski/ – /p`oisse/, /p`oi,sse/, /k`uuski/, /k`uus,ki/
poisse 'boys', part.pl., *poisse* 'buoy', illat.sg., *kuuski* 'spruces', part.pl., *kuuski* 'even six', nom.sg. + particle *-ki*
/ritva/, /rittva/, /rit:tva/ – /r´itva/, /r´ittva/, /r`ittva/
ridva 'rod', gen.sg., *Ritva* 'Ritva', a name, nom.sg., *ritva* 'rod', part.sg.
/sii:nne/, /kaa:rtu/, /pea:lse/ – /s`iinne/, /k`aartu/, /p`ealse/
siinne 'of this place', nom.sg., *kaardu* 'bent, warped', adv., *pealse* 'situated on something', gen.sg.
/maa:kki/, /mak:kki/, /maak:kki/ – /m`aak,ki/, /m`akkki/, /m`aakkki/
maagki 'even the magician', nom.sg. + particle *-ki*, *makkki* 'even the tape recorder', nom.sg. + particle *-ki*, *maakki* 'even the ore', nom.sg. + particle *-ki*

The main shortcoming of the system with colon is that only the A3 foot is designated. No problems arise in one-foot words because the difference between A1 and A2 is expressed merely by segmental duration and a special marking is not needed. But in longer words, if A1 and A2 are concerned, it is impossible to determine whether the utterance is a two- or three-footed word. However, in many cases the parsing can be automatically predicted (primacy of disyllabic foot division, broken by stressed suffixes

and long syllables, e.g. *kava-lama-leki* 'even to the slyer', allat.sg., *valet-tatteki* 'you even lie', 2.pl.pres., *kolet-tuttet-taki* 'even without monstrous', abess.pl., and *varesek-kene* 'little crow', nom.sg., *paranta-vatte* 'repairing', pres.participle in gen.pl., *paranta-mat:ta* 'unimproved; without improving', abess.sg., *vastas-tikkusta-mattut-tele* 'to the events that cannot be in reciprocal relations', allat.pl., etc.). But there are a lot of examples where the automatic rules do not work (e.g. *kavala-maleki* 'even to a slyer', allat.sg.). Although the phonological weight of parsing variability (dialectal? idio-syncratic?) is very small, it would be appropriate for the description of speech rhythm and the shifts within it to mark divisions caused by secondary stresses. For this purpose the system using acute and gravis is more pertinent (e.g. /k´avalam´atte/ 'slyer', gen.pl., /k´aval´amal´eki/ 'even to the slyer', allat.sg., /k´avalam´aleki/ 'even to the slyer', allat.sg., /t`öostust`esseki/ 'even to the industries', illat.pl., /p´attustam`atta/ 'without sinning', abess.sg.). In spite of the redundancy, it could be reasonable to unite both transcription modes in studying the word rhythm.

4. Results and Discussion

4.1. Experiment 1. In this experiment we studied whether listeners are able to identify quantity degrees (Q's) of phonemes in the first syllable of disyllabic words on the basis of the comparison of VC durations without receiving information about the vowel of the second syllable.

In the base words presented in Tables 4 and 5 the final part of the word was deleted by a window of different width in the way that the VC sequence in the initial part of the word was preserved (below in the text the point of deleting the final part of the word is marked with a slash). The initial parts of the words were presented to the listeners in random order. Two trained workers of the Laboratory of Phonetics and Speech Technology (Tallinn) were used as listeners.

After the listeners had marked short vowels and consonants as Q1 but had not been able to identify long sounds as Q2 and Q3 phonemes the task was changed. Listeners were asked to note down whether the vowel and consonant in the item was short or long. In the third session the listeners were asked to determine the syllabic quantity degree (accent) of each CVC token.

In the first half of the words */pal|at/* 'pieces', nom.pl., */pal|att/* 'hospital ward', nom.sg., */sat|a/* 'hundred', nom.sg., */tap|a/* 'padlock', nom.sg., */sam|a/* 'the same', nom.sg., the vowel as well as the consonant were perceived as short. In the first half of the words */saat|a/* 'send', 2.sg.imperat., and */saa:t|a/* 'to get', da-infinitive, a long vowel and a short consonant were perceived (as in the monosyllabic word *saad* 'haycock', nom.sg.). The first stimulus was perceived as a question or surprise (rising F0 of A2 base word) and the second as an answer (falling F0 of A3 base word). The same F0 difference is heard also in the first halves of the words */laut|a/* 'cowhouse', gen.sg., and */lau:t|a/* 'table', part.sg. In spite of these distinctions it was not possible to identify accents.

In the base words with long intervocalic consonants or clusters the duration perception depended on the width of the given window. In words */sam|ma/* 'sturgeon', gen.sg., */sam|pa/* 'samba', nom.sg., */sam|ppa/* 'champagne', nom.sg., */kam|p:pa/* 'troop', part.sg., a short vowel and a short consonant were heard in the first part

(separated by a slash) of the word. Also in the word */tap|pa/* 'kill', 2.sg.imperat., a short vowel and consonant were heard (as in *tab*) when the plosive occlusion has been shortened so that the duration of /p/ was equal to the duration of the first component of the geminate presented in Table 5. In the first parts of the words */sam:|ma/* 'the same', illat.sg., */sam:|pa/* 'column', gen.sg., */kamp:|pa/* 'troop', part.sg., */tap:|pa/* 'to kill', da-infinitive, a short vowel and a long consonant were perceived (as in words *samm* 'step', nom.sg., *kamp* 'troop', nom.sg., *tapp* 'transport of convicts', nom.sg.). The listeners perceived a short vowel and a long consonant also when the first part of the word up to V2 was made available to them: */samm|a/* 'sturgeon', gen.sg. – */sam:m|a/* 'the same', illat.sg., as *samm* 'step', nom.sg., but in interrogative and affirmative intonation respectively; */samp|a/* 'samba', nom.sg. – */sam:p|a/* 'column', gen.sg., like *samb* 'sturgeon', nom.sg.; */sampp|a/* 'champagne', nom.sg. – */kamp:p|a/* 'troop', part.sg., like *kamp* 'troop', nom.sg.; */tap:p|a/* 'to kill', da-infinitive, like *tapp* 'transport of convicts', nom.sg. The first part of the base word */tapp|a/* 'kill', 2.sg.imperat. was perceived as *taap* due to the relatively long V1 (see Table 5), i.e. V1 is relatively long if compared to the following long plosive but it is not long when the listener also gets information about V2.

Comparing the V+C durations, the short and long sounds can be identified (see Fig.1: the first stage of processing — Speech flow without V2), but it was impossible to ascertain in which of the three segmental duration degrees was the respective sound. Accents can be identified only when the listeners have the information about V2, i.e. not before the speech flow has been divided into two syllables (see Fig. 1: the second stage of processing — Speech flow with V2). However, listeners mentioned that the whole sequence in the CV(V)C(C)| window-grid seemed to them sometimes (i.e. in the cases when the stimuli had been generated from an A3 word) more likely to have a special peakedness but they could not associate these observations with syllabic Q's.

It is necessary to add that the syllable boundary is not perceived unless the second syllable nucleus (or at least a part of it) has been heard. Only then can syllabification take place.

Considering that the speech flow arrives to the ear linearly, we see that at the first stage it is possible to identify only short and long segments, and after syllabification the accents. In cases where the signal in the first stage of processing is directed to 'short segments', but later, in the second stage of processing, after V2 has been delivered and if it proves that the duration ratio is not in accordance with the A1 ratio V1<V2, this signal is sent as 'fast-rate stimulus' to 'long segments' where it is classified according to the parameters of A2 and A3 (see Fig. 1; e.g. stimulus 1e in Experiment 3).

We succeeded in showing the existence of only two (short vs. long), not three phonological quantity degrees on segmental and syllabic levels.

4.2. Experiment 2. In this experiment we study the question whether the information located in the end of the stressed syllable and the vowel of the unstressed syllable is sufficient for the listener to identify the accents.

In A2 and A3 base words presented in Tables 4 and 5 the first part of the word was deleted, i.e. the onset and a short V1 of the first syllable, while in the diphthongs the first component and in long vowels the segment with the length of a short vowel were

Table 4. Measurement data of the base words *palad*[1] – *palat*[2]; *sada*[3] – *saada*[4] – *saada*[5]; *lauda*[6] – *lauda*[7]. Male speaker.

Base word	Duration of segments, ms						Duration ratios		Values of the F0 contours in $(V_1V_2)1$, Hz				Values of the F0 contours in V2, Hz		Average intensity, dB	
	C1	$(V_1V_2)1$	C2	V2	$(C_1C_2)3$	Total	S1:S2**	M2:M3***	Initial F0	F0 peak	Final F0	F0 peak, % of $(V_1V_2)1$ duration	Initial F0	Final F0	$(V_1V_2)1$	V2
palat*	107	107	54	270	143	681	0.4	–	115	130	129	67	121	80	70	66
palatt	165	89	39	131	287	711	0.7	–	116	126	126	85	115	85	69	67
sata	136	126	73	233		568	0.5	–	143	176	176	79	165	105	74	75
saata	133	221	72	178		604	1.2	0.5	143	174	168	50	129	105	76	75
		126 95														
saa:ta	137	262	80	105		584	2.5	1.3	157	167	121	24	116	108	77	72
		126 136														
lauta	77	229	65	184		555	1.2	0.5	100	151	136	45	96	80	72	69
		142 87														
lau:ta	87	338	95	81		601	4.2	2.1	101	133	85	26	82	82	71	67
		164 174														

* Base words are presented here and henceforth in the phonological transcription, pertinent to our interpretation of the Estonian accents.
** S1 – nucleus of the first syllable (onset excluded), S2 – nucleus of the second syllable (onset and coda excluded).
*** M2 – the second mora (the second component of a geminate vowel or a diphthong in the first syllable), M3 – the third mora (a vowel of the second syllable).
palad[1] 'pieces', nom.pl.; *palat*[2] 'hospital ward', nom.sg.; *sada*[3] 'hundred', nom.sg.; *saada*[4] 'send', 2.sg.imperat.; *saada*[5] 'to get', da-infinitive; *lauda*[6] 'cowhouse', gen.sg.; *lauda*[7] 'table', part.sg.

Table 5. Measurement data of the base words *taba*[1] – *tapa*[2] – *tappa*[3]; *sama*[4] – *samma*[5] – *samma*[6]; *samba*[7] – *samba*[8]; *tahma*[9] – *tahma*[10]; *sampa*[11] – *kampa*[12]. Male speaker.

Base word	Duration of segments, ms					Duration ratios		Values of the F0 contours in V1, Hz				Values of the F0 contours in consonants, Hz		Values of the F0 contours in V2, Hz		Average intensity, dB	
	C1	V1	$(C_1C_2)2$; $C2(C_1C_2)3$	V2	Total	S1:S2*	M2:M3**	Initial F0	F0 peak	Final F0	F0 peak, % of V1 duration	Initial F0	Final F0	Initial F0	Final F0	V1	V2
tapa		120	74	242	436	0.5	–	146	174	172	77			155	98	74	73
tappa		132	178	163	473	1.4	0.6	136	174	170	76			131	110	72	72
			104 74														
tap:pa		95	277	87	459	3.4	2.3	145	155	149	63			123	107	73	70
			203 74														
sama	145	116	59	167	487	0.7	–	134	151	144	80			131	99	74	72
samma	123	142	139	170	574	1.3	0.5	133	167	167	82			125	103	72	71
			80 59														
sam:ma	118	108	226	98	550	2.8	1.7	149	154	151	57			100	86	74	64
			167 59														
sampa	112	128	135	138	513	1.5	0.5	122	157	157	100	157	132	129	103	70	71
			75 60														
sam:pa	112	103	166	85	466	2.5	1.3	128	138	137	66	137	106	106	98	73	68
			113 53														
tahma		122	208	182	512	1.3	0.6	140	178	172	84	148		121	111	75	70
			118 90														
tah:ma		117	293	77	487	3.9	2.4	143	157	154	56	110		110	94	72	63
			187 106														
samppa		117	73 104	143	437	1.7	0.4	119	155	155	100	155	147	140	96	73	72
			51 53														
kamp:pa		97	96 166	66	425	4.6	1.7	126	147	147	100	147	121	105	99	73	68
			113 53														

*S1 – rhyme of the first syllable (onset excluded), S2 – nucleus of the second syllable (onset excluded).
** M2 – the second mora (coda peak), M3 – the third mora (a vowel of the second syllable).
taba[1] 'padlock', nom.sg.; *tapa*[2] 'kill', 2.sg.imperat.; *tappa*[3] 'to kill', da-infinitive; *sama*[4] 'the same', nom.sg.; *samma*[5] 'sturgeon', gen.sg.; *samma*[6] 'the same', illat.sg.; *samba*[7] 'samba', nom.sg.; *samba*[8] 'column', gen.sg.; *tahma*[9] 'soot', gen.sg.; *tahma*[10] 'soot', part.sg.; *sampa*[11] 'champagne', nom.sg., fam.; *kampa*[12] 'troop', part.sg.

deleted. In order to make the plosives heard as a geminate in the words */samppa/* 'champagne', nom.sg. – */kamp:pa/* 'troop', part.sg., the 1-2 last periods of /m/ were not deleted; in the pair */tappa/* 'kill', 2.sg.imperat. – */tap:pa/* 'to kill', da-infinitive, the same effect can be created by not deleting the ca 10-15 ms preaspiration preceding the plosive occlusion. When the consonant preceding V2 is two times longer than the short consonant, the syllabification is clearly perceived: we hear a geminate divided by the syllable boundary even without V1. These minimal units that carry accent features were presented to the listeners (see Experiment 1) in random order and they were asked to identify the units as A2 or A3.

The parts *mma – m:ma* (< */samma/* 'sturgeon', gen.sg. – */sam:ma/* 'the same', illat.sg.), *mpa – m:pa* (< */sampa/* 'samba', nom.sg. – */sam:pa/* 'column', gen.sg.), *hma – h:ma* (< */tahma/* 'soot', gen.sg. – */tah:ma/* 'soot', part.sg.), *^{m}ppa – ^{m}p:pa* (< */samppa/* 'champagne', nom.sg.– */kamp:pa/* 'troop', part.sg.), *^{h}ppa – ^{h}p:pa* (< */tappa/* 'kill', 2.sg.imperat. – */tap:pa/* 'to kill', da-infinitive), separated from the base words, were perceived as A2 and A3 units like the respective original words. The pairs separated from a long open first syllable *ata – a:ta* (< */saata/* 'send', 2.sg.imperat. – */saa:ta/* 'to get', da-infinitive) and *uta – u:ta* (< */lauta/* 'cowhouse', gen.sg. – */lau:ta/* 'table', part.sg.) were perceived as fast-rate A2 (?) and A3 units. Curtailed /ata/ and /uta/ units were perceived as A1 units rather than fast-rate A2 units. The listeners' opinions are based on the fact that the ratio of the curtailed units is in accordance with the A1 ratio V1<V2 (see Table 4) and that in the first stage of processing a short vowel + a short consonant can be perceived (see Figure 1).

The results of this experiment support the statement that the cues characterizing accents are concentrated in the end of the stressed syllable and in the vowel of the unstressed syllable (A2: high F0, balanced energy; A3: low F0, unbalanced energy, see data in Tables 4 and 5). Also the duration ratios of these parts in A2 and A3 are diametrical (see M2:M3, Tables 4 and 5), thus, according to the data of Fox and Lehiste (1987, 1989) these ratios should be discriminated. Certainly, in the process of perception the listener cannot skip the short first-syllable vowel or the first component of the long vowel (i.e. the first mora). But the compact concentration of the most important information on both sides of the syllable boundary should make the identification of accents easier. At the same time the first mora carries information about whether the following item is short or long (cf. the first-level processing, Fig.1).

For the identification of accents, not all phonemes in a stress foot (which can be a trisyllabic unit in Estonian) — and as we can see now — even in the stressed and the following unstressed syllable, are of the same importance. Syllable onsets and the whole third syllable (if the word happens to contain one) do not belong to the domain of accents. The main region, essential for the perception of accent types in the stress foot, consists of the rhyme of the first syllable (or minimally its final part, i.e. the second half of the long nucleus or the coda peak), and of the nucleus of the unstressed second syllable. The concentration of the perceptually relevant information around the syllable boundary (as is shown in Experiment 2) creates a perceptual impression that A3 has always a certain kind of peakedness whereas A2 has not.

However, it is convenient to separate two different aspects of the problems studied. Identifying the speech segments that carry sufficient information for the perception of

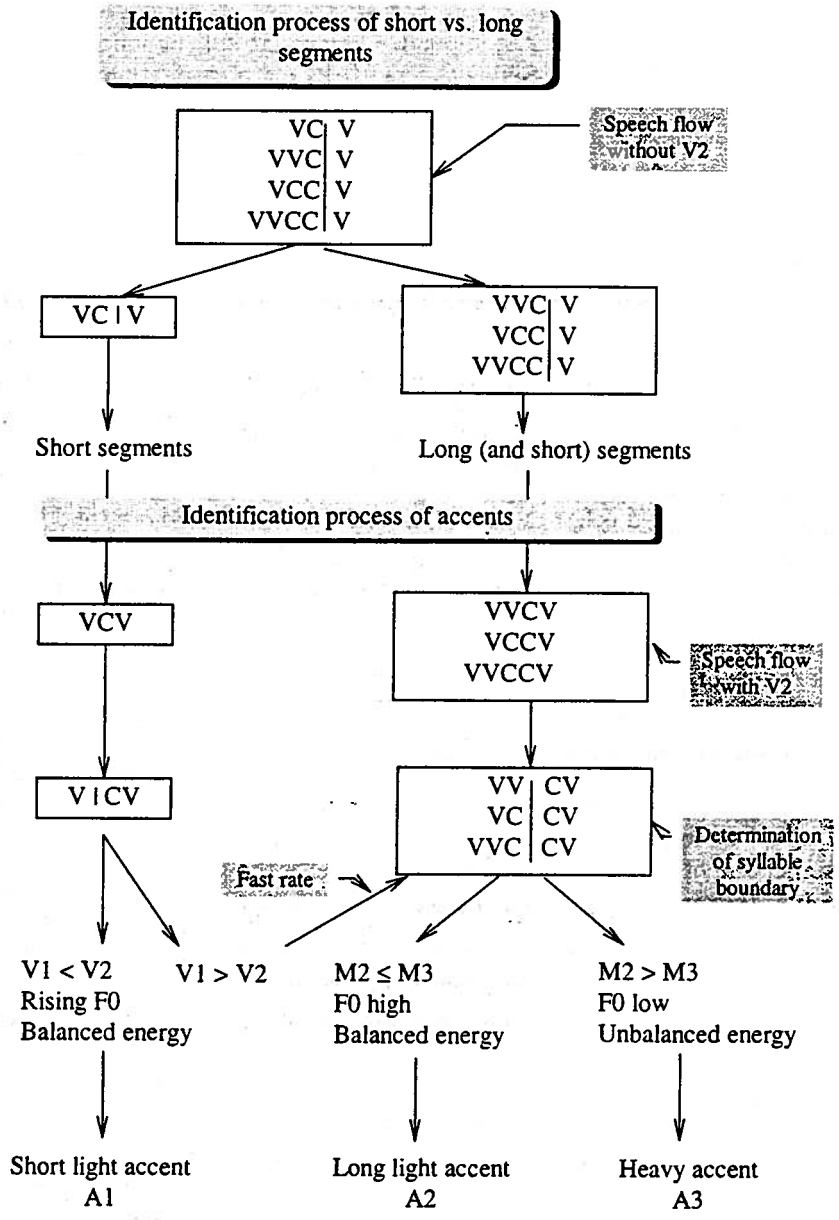

Figure 1. A schematic diagram of the probable two-level identification process of Estonian short/long segments and foot accents.

accents, is one aspect; the other concerns the units of temporal regulation. Two temporal regulation mechanisms seem to be operating in a stress foot (up to a trisyllabic unit). The higher-level temporal regulation, determined by the stress for the whole stress foot, is expressed in an isochronic trend according to which adding sounds to a foot decreases the duration of constituent segments in the foot, but does not affect its accent. In the stress foot a smaller isochronic unit (sometimes coinciding with the stress foot) can be observed. This lower-level temporal regulation mechanism, subordinated to the whole stress foot, characterizes isochrony of the disyllabic accent foot (see Table 6).

Accent isochrony functions even in the case of syllable-counting strategy when the stress pattern of the word has changed (cf. Hint 1978). As is known, due to affixation, the natural speech rhythm is distorted and the so-called morphologically bound secondary stress has arisen (Hint 1973). Adjacent stressed syllables, which are unnatural from the point of view of stress rhythm, have appeared in Estonian (e.g. *'met:s,nik:k* 'forester', nom.sg., *'met:s,nik:kute* 'forester', gen.pl.). This can happen when the first of the two successive syllables bears the heavy accent. In contemporary colloquial speech the second-syllable secondary stress has disappeared and is disappearing even from those suffix morphemes which have always required a secondary stress. As a result, the tendency toward restoring the natural stress pattern (whose leading principle forbids adjacent stressed syllables) has already created new alternation and declension types (e.g. *'metsnikk* or *'met:snikk* 'forester', nom.sg.; *'metsnik,kutte* or *'met:snik,kutte* 'forester', gen.pl.). It is also supposed that in these examples which have undergone the stress pattern shift, the rhyme of the first syllable and the nucleus of the second syllable function as a unit of accent isochrony.

Several works on moraic treatment of Estonian accents should be mentioned (e.g. Hint 1978; Prince 1980; Viik 1985; Hayes 1989; but see also Viitso 1982; Lehiste 1990). According to Trubetzkoy (1939) a mora analysis is motivated when (a) long vowels are treated in the same way as diphthongs in the functioning of the system; (b) a long vowel is comparable to two short ones or to a short vowel + a consonant in the system of accents; (c) long vowels make a phonological distinction between two types of accents. In mora-counting languages a long syllable nucleus can be better analyzed as geminated, and in these languages the smallest prosodic unit does not always coincide with the syllable. All these features characterizing mora-counting languages occur also in Estonian.

We use the mora as a minimal abstract unit of temporal regulation. It is assumed that syllables may contain either one or two moras (see below: Figures 2 and 3). Data in Table 6 provide evidence for the isochronic nature of the accent feet. Total durations of the accent feet do not differ essentially. The same tendency to equalize total durations of accent feet can be noticed even in the case of more complicated foot structures (e.g. the pairs /taata/ – /taa:ta/ and /taatta/ – /taat:ta/ or /taartta/ – /taart:ta/ etc. in Eek 1975: Table 3; we do not present these re-computations here). As a result of intra- and intersyllabic temporal compensation the total duration of disyllabic accents is consistent with our trimoraic treatment presented in Table 6. Following the criterion that syllables can be maximally bimoraic, and that complicated structures have more or less the same total accent foot duration as simpler structures (e.g. CVV|CV or

Table 6. Durations in milliseconds of differently selected units of accent isochrony (based on the data in Tables 4 and 5).

Words and moras	Trimoraic accent feet		Two-mora units consisting of the 2nd and 3rd mora	
	Duration of the foot	Duration per mora	Duration of the unit	Duration per mora
1. 2. 3.				
p a l a t	377	126		
p a l aa t	364	121		
s a m a	283	94		
s a t a	359	132		
t a p a	362	121		
Average	349	119		
1. 2. 3.				
s a m a	392	131	250	125
s a m a	373	124	265	133
s a p a	341	114	213	107
s a p a	301	100	198	99
t a m a	422	141	300	150
t a m a	381	127	264	132
t a p a	399	133	267	134
t a p a	385	128	290	145
s a t a	399	133	273	137
s a t a	367	122	241	121
l a t a	413	138	271	136
l a t a	419	140	255	128
Average	383	128	257	129

CVC|CV), we intend to use the mora-splitting procedure (see Maddieson 1993, where multiple linking of segments to moras is interpreted as sharing of this abstract durational unit). Splitting moras is a convenient procedure for the description of intrasyllabic compensation phenomena and it avoids the need to define tri- or quadri-moraic syllables. Moreover, accent-dependent intersyllabic temporal reorganization in accent feet (e.g. the well-known inversely proportional duration relations) can also be interpreted as mora-sharing. It appears particularly evident if one compares relations of the second and third moras in the accent foot, i.e. the moras of the segments around the syllable boundary whose total durations have a trend to isochrony (see the right-hand data in Table 6; cf. also the base data in Tables 4 and 5). The impressionistic percept — peakedness of A3 — is based, among other cues (see Experiment 3), also on the temporal aspect: a part of the duration of the third mora is drawn to the second mora, enhancing lack of balance of the A3 foot whereas a part of the second mora of A2 is pushed to the third mora segment, strengthening balancedness of the A2 foot.

4.3. Experiment 3. The task of this experiment was to examine under which conditions it could be possible to change an A3 word to a natural-sounding A2 word and vice versa.

Changed parameters of the base words, presented in Tables 4 and 5, are durations of the second and third moras, the intensity of the third mora (V2) and the fundamental frequency of the first syllable. The modifications of the base words were presented in

random order to two listeners (see Experiment 1) whose task was to identify the accent of the stimuli. Perception results are presented by base words. We treat only unanimous judgements.

(1) *A vowel-peaked type.* Base word /saa:ta/ 'to get', da-infinitive (A3; S1:S2=2.5; M2:M3=1.3; low F0; unbalanced energy; see Table 4).

(a) The A3 base word is perceived as A3 word even after the original falling F0 contour is changed into a rising one, while all other parameters are left unchanged. Also in A2 and A1 words, changing the F0 contour into a falling one and leaving all other parameters unchanged does not lead to a change in accent perception.

(b) When V1 of an A3 base word is shortened and the qualitatively reduced weak V2 is lengthened up to the corresponding durations of an A2 word (S1:S2=1.3; M2:M3=0.5), while F0 remains unchanged (falling), the stimulus is still perceived as an A3 word. Both the falling F0 and the weak and reduced V2 (although lengthened) work in favour of A3. A rising F0 turns the responses in favour of A2, in spite of the fact that the difference in V1 and V2 amplitudes was retained. The influence of the fundamental frequency is greater, inhibiting the countereffect of the relatively weak V2 (which is unnatural for an A2 word). As we mentioned in 1a, the A2 base word is perceived as an A2 word even with a falling F0. But then the equality of V1 and V2 amplitudes was preserved. Thus the duration ratio and balanced energy together outweigh the F0 influence in an A2 base word. But the same parameters in an A3 base word modification, made to correspond to the respective parameter values of the A2 base word, do not lead to the recognition of A2; the rising F0 contour is necessary as well. To overcome the countereffect of A3 in a vowel-peaked type, the duration ratio and the fundamental frequency must work together.

(c) It is possible to get a natural-sounding A1 word from an A3 base word only if V1 is made extra short (the only case where the absolute duration of V1 is of importance) and the duration ratio is extremely small (V1:V2≤0.5); then the characteristics of F0 contour and V2 amplitude do not influence the judgements.

Base word /saata/ 'send', 2.sg.imperat. (A2; S1:S2=1.2; M2:M3=0.5; high F0; balanced energy).

(d) If the V1 duration of the A2 base word /saata/ is shortened to the V1 of an A1 word (duration ratio V1:V2=0.7, characteristic of A1), then the F0 contour has no influence on judgements: both rising and falling F0 are equally well perceived as A1.

(e) When the duration of V1 of the A2 base word was made equal to that of V1 of an A1 word (as in 1d) and at the same time V2 was shortened creating a ratio characteristic of A2 (V1:V2=1.2), while the rising F0 contour was retained, then the stimulus was perceived as a fast-rate A2. The absolute duration turned out not to be essential.

This case needs some explanation. Processing the stimulus 1e could be as follows. On the short/long identification level the first part of /sat|a/ (see Fig. 1) is classified as 'short vowel' + 'short consonant'; after hearing V2, the listener recognizes that the stimulus does not belong to the A1 category because V1>V2, and the stimulus is labelled 'fast-rate' stimulus and directed to the category 'long segment' where the duration ratio corresponds to A2 (S1:S2=1.2, or M2:M3=0.5). When V2 is still

shortened (S1:S2=2.5, or M2:M3=1.3) and the F0 contour is turned into a falling one, we get a fast-rate A3 word.

As an Estonian listener, according to Fox and Lehiste (1987, 1989) does not distinguish the V1:V2 ratios of A2 and A3 (in both accents V1>V2), there is a possibility in our processing mechanism that after the sound has been labelled as 'long', it will be divided into two parts between the first and the second mora. That is the reason why we use the ratio M2:M3; this ratio expresses the relations between the durations of V1 and V2 in A2 and A3 words.

(f) If V2 is shortened in the A2 base word to the extent that the duration ratio is already characteristic to an A3 word (S1:S2=2.6, or M2:M3=1.1), but F0 is retained rising, then the stimulus is still perceived as an A2 word. If V2 is shortened even more, so that the duration ratio conspicuously expresses an A3 word (S1:S2=3.2, or M2:M3=1.4) and if, in addition, the F0 contour is changed to falling, only then is the word perceived as A3. Thus, all parameters, except V2 amplitude, have to be in use. Obviously the longer duration in the first syllable creates the impression of unbalanced energy (the first syllable is stronger than the second one), and there is no need to reduce the amplitude of V2 for the identification of A3 in the vowel-peaked type. This corroborates the findings of Lehiste and Fox (1992) that in the perception of prominence Estonian listeners are more responsive to duration than to amplitude cues.

(2) *Consonant-peaked types*. Base word /tah:ma/ 'soot', part.sg. (A3; S1:S2=3.9; M2:M3=2.4; unbalanced energy).

(a) When /h/ in the A3 base word was shortened up to the absolute value of /h/ in the A2 word, but the duration ratio typical to A3 was retained (S1:S2=3.0, or M2:M3=1.5), the responses were A3.

When the duration of a weak V2 of the same stimulus was lengthened so that a typical A2 duration ratio was achieved (S1:S2=1.5, or M2:M3=0.8), the responses were still A3.

Although V2 was lengthened even more, up to an A2 duration ratio S1:S2=1.0, or M2:M3=0.5, the stimulus was still perceived as A3. The impression of unbalanced energy, evoked by the cooccurrence of the strong (although shortened) friction of /h/ in the final part of the first syllable of the A3 base word and the weak (although lengthened) V2, eliminated the relative importance of duration ratios.

(b) Increasing the amplitude of V2 of the stimulus with the most effective A2 duration ratio (2a) resulted in an unnatural A3. In addition to everything that had been done already, weakening of the /h/ in the first syllable was also needed for perceiving a natural A2 word. To neutralize unbalanced energy in the voiceless part of a consonant-peaked A3 word, changing of the duration ratio only was not sufficient; the amplitude also plays a role for balancing energy. This is a remarkable difference between vowel-peaked and consonant-peaked types.

5. Conclusions and Generalizations for a Conceptual Frame[1]

The main principles of a conceptual framework for the description of Estonian word prosody are schematically presented in Figures 2 and 3. Below we add some concluding remarks and explanations.

We have not succeeded in showing three phonological quantity degrees on segmental and syllabic levels.

All sounds have two, short and long, distinctive duration degrees on the segmental level. The long degree is represented by a unit consisting of two short phonemes. Comparing the durations of consecutive vowels and consonants in V(V)C(C)| units during the first stage of signal processing (Experiment 1), only short and long sounds can be identified. In this stage of the incoming speech, it is impossible to ascertain three distinctive duration degrees of the segments by perception.

Syllables are short and long (or light and heavy; depending on foot branching — weak and strong). Segmental and syllabic length are identified synchronously, shifting a V(V)C(C)| window-grid continuously with the incoming speech up to the end of a word. A syllable is long when the window-grid captures a long vowel, a diphthong, a long consonant, a consonant cluster, or long vowel and long consonant simultaneously. Estonian belongs to the group of languages characterized as WEIGHT-by-POSITION (Hayes 1989) languages. According to this typological feature Estonian listeners transfer (after they have perceived a syllable boundary) a short single consonant and the end part of a geminate or a consonant cluster to the onset of the next syllable if there happens to be one in the word. Therefore, it is quite plausible that a coda consonant is monomoraic and a short intervocalic consonant is non-moraic. A short word-final consonant is also non-moraic due to its potential to act as an onset when a vowel is added to the word-final consonant in the word paradigm (cf. e.g. the so-called floating onset in /kalat/ 'fishes', nom.pl., and /kalate/ 'fishes', gen.pl.). Short vowels are monomoraic; long vowels and diphthongs are bimoraic. If a phonologically short vowel has been perceived as phonetically long during the first stage of signal processing (e.g. the second-syllable vowel in /vilel/ 'whistle', adess.sg.; cf. /filee:l/ 'fillet', adess.sg.), the correction 'phonetically long = phonologically short' must be made in the provisional judgement after stress has been perceived on the preceding syllable. Due to the neutralization of the short/long vowel contrast in the unstressed second syllable of A1 (and A2), the listeners' phonological competence acts there as a masker (see Table 6 and Fig. 3).

Dictated by SYLLABLE MAXIMALITY, a long syllable is maximally bimoraic and a short syllable is monomoraic (below we will explain how to avoid inadmissible tri- or quadrimoraic syllables).

The mora is a minimal abstract unit of temporal organization in the minimal distinctive prosodic unit. And even in this case, a mora is a meaningful and useful construct only if one admits that the mora cannot be linked to all phonemes in the distinctive prosodic sequence (e.g. syllable onsets in Estonian are extramoraic). The smallest prosodic unit in Estonian does not coincide with the syllable. We argue that disyllabic accents (accent feet) are minimal distinctive prosodic units in Estonian. The mora is an appropriate unit substantiating temporal compensation in isochronic accents.

Accents (traditionally called syllabic quantity degrees) can be identified only when information about V2 has also been delivered to listeners (Fig. 1). This points to the accentual character of quantity degrees. We have discarded the concept of three syllabic quantity degrees and substituted the notion of three distinctive accents. The main region of the stress foot that is essential for the perception of accent types consists

of the stressed syllable rhyme (for A2 and A3: minimally its final part, i.e. the final half of the long nucleus or the coda peak) and of the nucleus of the following unstressed syllable. The concentration of perceptually relevant information around the syllable boundary, as is shown in Experiment 2, creates a perceptual impression that A3 has always a certain kind of peakedness whereas A2 has not. All A3 feet are either vowel-peaked or consonant-peaked, other possibilities are lacking. Peakedness is an impressionistic percept based on the cooperation of several cues; the potential lengthening of the peaked phoneme represents a temporal cue. Although the peaked segment is usually longer than the average mora of the accent foot (cf. Tables 4, 5 and 6) and the peakedness itself is perceivable, listeners cannot identify these syllables as being in A3. In monosyllabic units, the A2/A3 contrast is neutralized. Besides segmental durations, listeners need additional information for the identification of accents (see Experiment 3) embedded in a wider context than is provided by the stressed syllable alone. In an unbalanced A3 disyllabic foot the perceptual effect of peakedness is strengthened against the background of the weakened, shortened, and qualitatively reduced second syllable vowel. The balance of A2 becomes conspicuos against the background of the lengthened and unreduced second syllable vowel.

In vowel-peaked feet the lengthened peaked vowel tends to press the following consonants out of the coda (cf. e.g. /soo:tne/ 'favorable', nom.sg., and /root:su/ 'rib of a leaf', part.sg.; /poi:sse/ 'buoy', illat.sg., and /pois:se/ 'boys', part.pl.). It seems that WEIGHT-by-POSITION does not work in vowel-peaked cases because the position of these consonants is ambiguous and their status in the foot becomes extramoraic (Fig. 3).

Accents, as the minimal distinctive prosodic units, are closely connected with the stress pattern of a word. Stress is a precondition for accents. The domain of accents is a stress foot, irrespective of whether the foot is primarily or secondarily stressed. Two temporal regulation mechanisms operate in the stress foot. The higher-level temporal regulation, determined by the stress for the whole stress foot (up to trisyllabic feet), is expressed in an isochronic trend according to which the adding of phonemes to the stress foot decreases the durations of the constituent segments in the foot, but does not affect accents. The lower-level temporal regulation mechanism, subordinated to the whole stress foot, characterizes isochrony of the accent foot (sometimes coinciding with the stress foot) whose segmental composition is predetermined by FOOT MINIMALITY. According to this criterion, a stress foot must be at least disyllabic. Therefore this is also valid for the accent foot. There is a remarkable difference between the two relatively independent types of temporal regulation. A stress-conditioned trend to isochrony influences absolute durations of the intrasyllabic constituents, whereas an accent-conditioned temporal regulation (e.g. the well-known inversely proportional duration relations between the first and second syllable of the foot) organizes the constituents of isochronic accent feet into relationships characteristic of individual accents (see Fig. 3).

Our data illustrate the isochronic nature of accent feet. Total durations of A2 and A3 feet (and also of A1 feet in the speech of speakers with a western dialectal background) do not differ essentially (Table 6). The total durations of complicated structures (e.g. /saatte/ 'you get', 2.pl.pres. – /saat:te/ 'sending', gen.sg., or /kaartte/

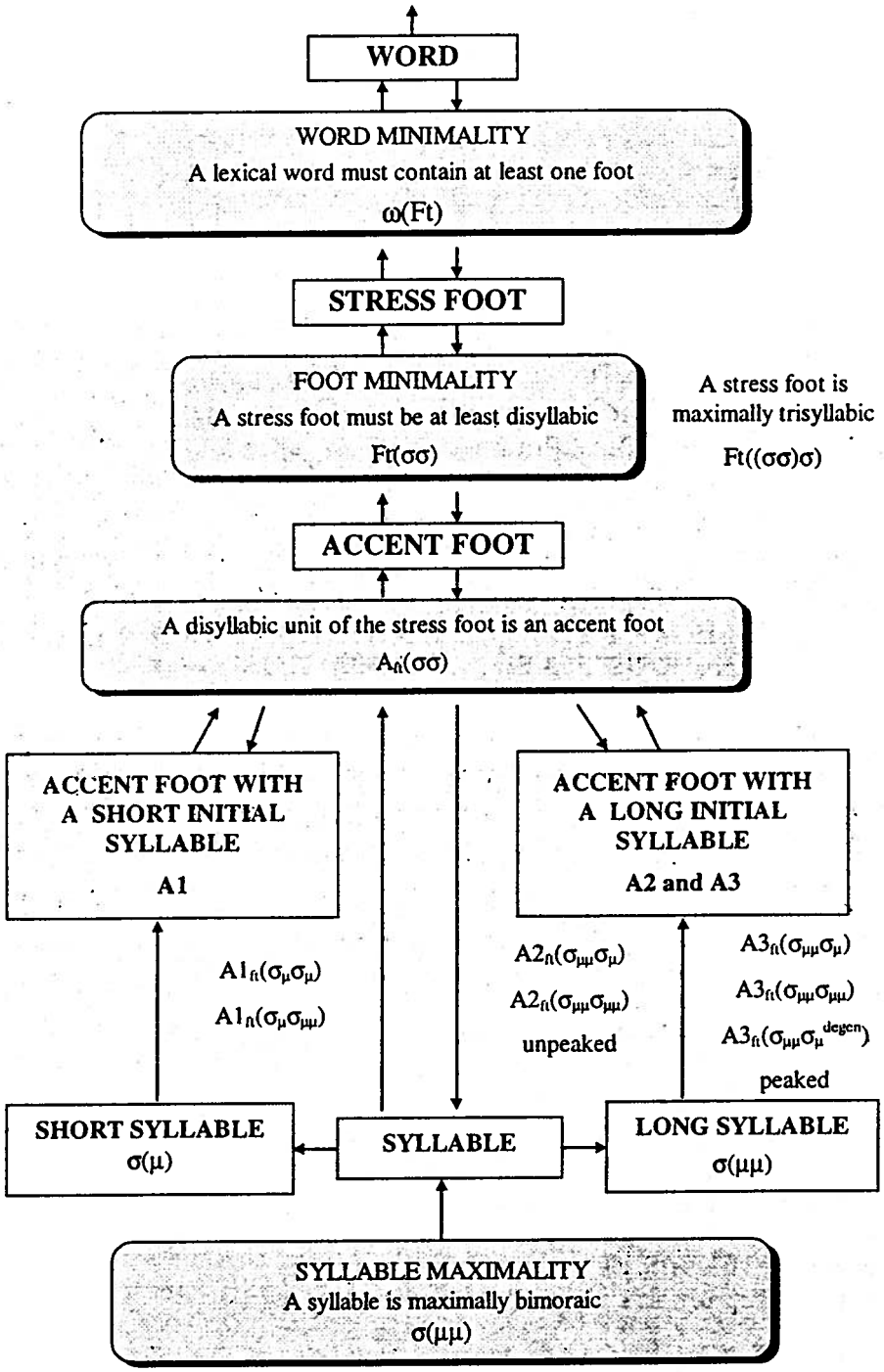

Figure 2. Prosodic hierarchy in Estonian words.

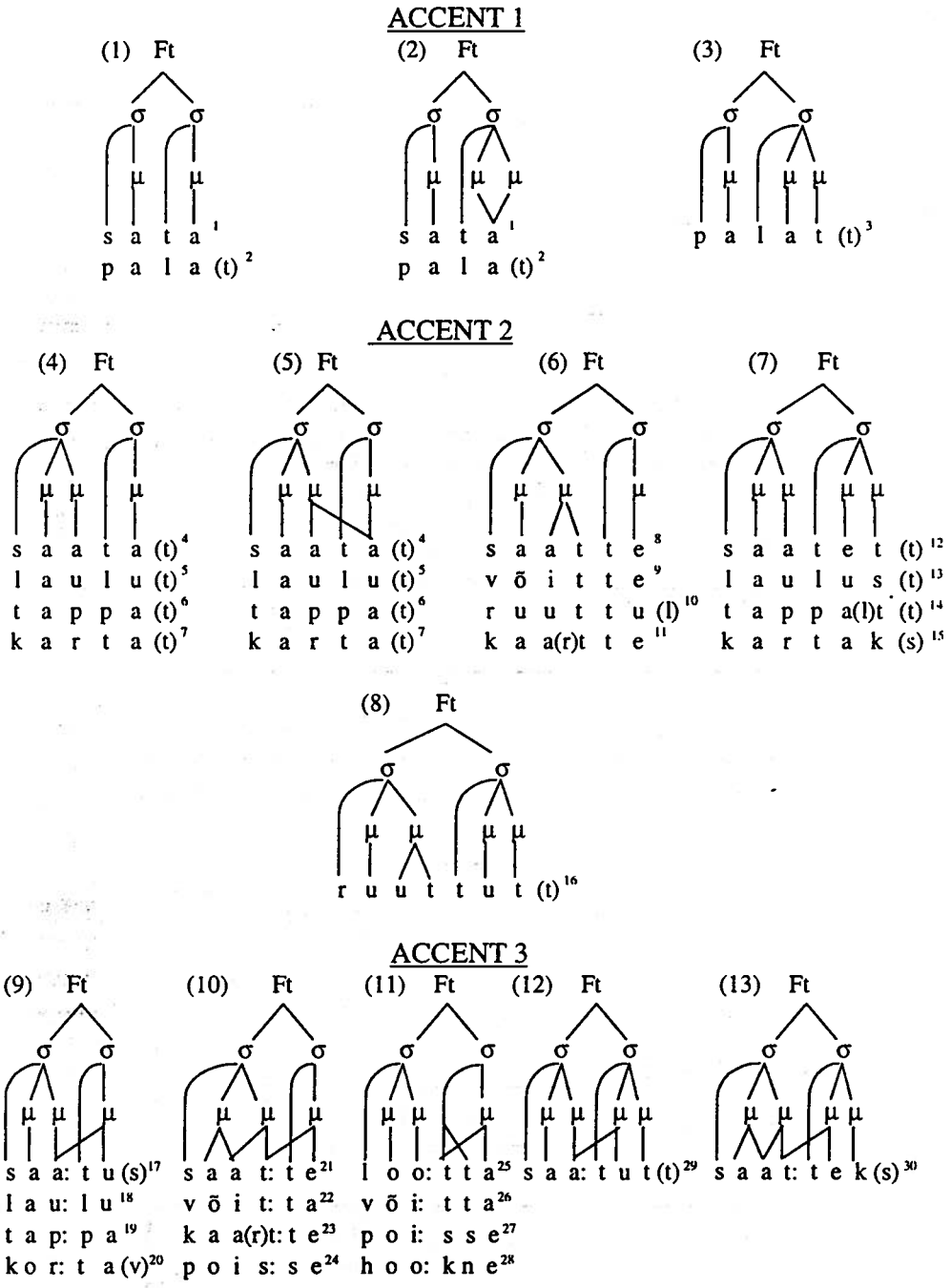

Figure 3. The main foot types of Estonian accents (about glosses see Note 2).

'windrows', gen.pl. – /kaart:te/ 'cards', part.pl.) are more or less the same when compared with the total foot durations of simpler structures (e.g. /saate/ 'sending', nom.sg. – /saa:te/ 'haycocks', part.pl., or /katte/ 'cover', nom.sg. – /kat:te/ 'cover', gen.sg.). The above-mentioned stress-conditioned trend to isochrony has decreased the absolute durations of the intrasyllabic segments in the feet with those complicated structures. This kind of temporal compensation gives us an opportunity to avoid more than two moras (see SYLLABLE MAXIMALITY) in syllable descriptions. Here we suggest to use the mora-splitting procedure (cf. Maddieson 1993, where multiple linking of segments to moras is interpreted as the sharing of this abstract durational unit). Moreover, accent-dependent intersyllabic temporal reorganization in accent feet can also be interpreted as mora-sharing: a part of the duration of the third mora is drawn to the second mora in an A3 foot, whereas a part of the second mora is pushed to the third mora in an A2 foot (Fig. 3).

FOOT MINIMALITY (i.e. a stress foot must be at least disyllabic) is based on the inherent property of stress. The stress foot is an inherently bipartite unit (Eek 1990: 253). The articulation of the speech flow can be regarded as a succession of energy pulses the general shape of which is physiologically determined by an alternation of tension and relaxation phases (here: by an alternation of the stressed and unstressed or the strong and weak syllables). A stress foot is maximally trisyllabic. Also in the latter case level pronunciation is avoided (Eek 1990: 259). Inherent binarity is noticeable in the frame of trisyllabic feet: the foot-branching creates the typical intra-foot alternation of strength degrees (e.g. Ft ((kava)la) 'sly', gen.sg., 132). Durational relations reflect this intra-foot parsing: the second syllable vowel is 'halflong' as usually in disyllabic A1 accent feet and the duration of the third syllable is similar to the duration of the short stressed syllable. The last syllable in the trisyllabic foot does not play a role in the identification of the accent type of the whole foot. This 'halfstrong' word-final syllable, if long (heavy), can acquire the secondary stress in correct pronunciation, irrespective of the accent type of the preceding foot and the length of the preceding syllable (e.g. parsing into two feet A1+A3: (pala)(tek:s) 'pieces', translat.pl., (palat)(tik:s) 'hospital ward', translat.sg.; A2+A3: (suure)(mak:s) 'bigger', translat.sg., (saatet)(tak:s) 'it would be sent', pres.condit.passive; A3+A3: (saa:tu)(tek:s) 'haycocks', translat.pl., (jälk:kus)(tek:s) 'disgust', translat.pl.).

The last examples draw attention to the cases where the A3 stressed syllable behaves like a disyllabic foot. It is well known that in hypercorrect pronunciation an A3 (monosyllabic) foot allows immediately following stress (e.g. (met:s)(nik:k) 'forester', nom.sg., (kau:)(kele) 'far away'), whereas the stressed syllable in A1 and A2 feet must be followed by at least one unstressed syllable. In accordance with the principles presented in Figure 2, this type of an A3 foot must be represented as a formal disyllable, i.e. a bimoraic syllable followed by a degenerate syllable: $A3_{ft}(\sigma\mu\mu+\sigma\mu^{degen})$. This solution, however, does not explain why the degenerate syllable notion cannot be applied to the bimoraic stressed syllable of an A2 foot. We must repeat here: even though the stressed syllable of an A2 foot is bimoraic, the whole foot must be balanced (unpeaked), and though the stressed syllable of an A3 foot is in principle bimoraic too, the whole foot must be unbalanced (peaked). The peakedness of A3 is attained by intersyllabic mora-sharing: a part of the second syllable mora is drawn to the second

mora of the first syllable. In the case of the postulated degenerate syllable this mora-sharing has resulted in full apocope as it has happened in language history. In the real disyllabic version the intersyllabic mora-sharing results in extra shortness of the second syllable nucleus. In both cases the same mechanism is used for expressing peakedness. But in the A2 foot the unpeakedness (balancedness) is attained by diametrically different intersyllabic mora-sharing: a part of the second mora is pushed to the third mora in the second syllable. This procedure results in halflongness of the second syllable nucleus. The characteristic trend to preserve balancedness and to avoid peakedness does not allow a secondary stress immediately following the stressed syllable in an A2 foot. In colloquial speech we meet A3 feet with a degenerate second syllable mainly in the case of separate words, e.g. (pat:t) 'sin' nom.sg., (kõrt:s) 'tavern', nom.sg., or in the word-final position, e.g. (kirjut)(tak:s) 'he would write', 3.sg.pres. condit., i.e. in the cases where the immediately following stress is not needed. At the same time, the first usage of the two possible variants, (kau:)(kele) or ((kau:ke)le) 'far away', (met:s)(nik:k) or (met:snikk) 'forester', nom.sg., is met only in hypercorrect or emphatic pronunciation. We suppose that the mentioned variability demonstrates the primacy of the stress rhythm patterning in colloquial speech trying to avoid two consecutive stressed syllables. Accent feet are subordinated to the higher level stress constraints.

References

Ariste, P. 1939. A quantitative language. Proceedings of the Third International Congress of Phonetic Sciences: 276–280. Ghent.

Ariste, P. 1953. Foneem eesti keeles. Eesti NSV Teaduste Akadeemia Toimetised, Ühiskonna-teaduste Seeria 3:357–367.

Eek, A. 1973. Observations in Estonian palatalization: An articulatory study. Estonian Papers in Phonetics, EPP 1973: 18–37. Tallinn.

Eek, A. 1974. Observations on the duration of some word structures: I. Estonian Papers in Phonetics, EPP 1974: 18–31.

Eek, A. 1975. Observations on the duration of some word structures: II. Estonian Papers in Phonetics, EPP 1975: 7–54.

Eek, A. 1976. Production of single and geminate labials on neuro-muscular level. 14. akustická konferencia. Akustika rečí a vnímanie zvuku, 1. časť: 30–33. Bratislava.

Eek, A. 1977. Experiments on the perception of some word series in Estonian. Estonian Papers in Phonetics, EPP 1977: 7–33.

Eek, A. 1980a. Estonian quantity: Notes on the perception of duration. Estonian Papers in Phonetics, EPP 1979: 5–30.

Eek, A. 1980b. Further information on the perception of Estonian quantity. Estonian Papers in Phonetics, EPP 1979: 31–57.

Eek, A. 1983. Kvantiteet ja rõhk eesti keeles (I): Fonoloogiliste tõlgenduste kriitikat. Keel ja Kirjandus 9:481–489, 10: 549–559.

Eek, A. 1986. Problems of Estonian word prosody. Estonian Papers in Phonetics, EPP 1984–1985: 13–66.

Eek, A. 1990. Units of temporal organization and word accents in Estonian. Linguistica Uralica 26 (4): 251–263.

Eek, A. 1994a. Studies on quantity and stress in Estonian. Dissertationes Philologiae Estonicae Universitatis Tartuensis 4. Tartu.

Eek, A. 1994b. Quantity and geminates. Paper read on a seminar at the Institute of Linguistics, University of Stockholm, June 8, 1994.

Eek, A., M. Remmel. 1975. Context, contacts and duration: Two results concerning temporal organization. Proceedings of the Speech Communication Seminar, Stockholm, August 1–3. 1974. Volume 2 (ed. G. Fant): Speech Production and Synthesis by Rules : 187–192. New York, London, Sydney, Toronto.

Engstrand, O., D. Krull. 1994. Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. Institute of Linguistics, University of Stockholm, PERILUS 18: 39–54.

Fox, R. A., I. Lehiste. 1987. Discrimination of duration ratios by native English and Estonian listeners. Journal of Phonetics 15: 349–363.

Fox, R. A., I. Lehiste. 1989. Discrimination of duration ratios in bisyllabic tokens by native English and Estonian listeners. Journal of Phonetics 17: 167–174.

Hallap, V. 1963. Fonoloogiline etüüd eesti keele väldete alalt. Nonaginta: 95–122. Tallinn: Emakeele Selts.

Harms, R.T. 1962. Estonian grammar. Indiana University Publications, Uralic and Altaic Series 12. Bloomington: Indiana University Press.

Harms, R.T. 1978. Some observations and hunches concerning Estonian prosody. Estonian Papers in Phonetics, EPP 1978: 31–34.

Hayes, B. 1989. Compensatory lengthening in moraic phonology. Linguistic Inquiry 20 (2): 253–306.

Hint, M. 1966. On the phonological transcription of overlength in Standard Estonian. Soviet Finno-Ugric Studies 2 (1): 23–36.

Hint, M. 1973. Eesti keele sõnafonoloogia. I. Rõhusüsteemi fonoloogia ja morfofonoloogia põhiprobleemid. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Hint, M. 1977. Ühesilbiliste sõnade vältest. Keel ja Kirjandus 4: 219–229, 5: 271–277.

Hint, M. 1978. Changes in the prosodical system of contemporary Estonian. Estonian Papers in Phonetics, EPP 1978: 39–43.

Hint, M. 1983. Viroko neljän kvantiteettiasteen sävelkieli? Virittäjä 2: 161–177, 3: 328–348. Helsinki.

Krull, D. 1991. Stability in some Estonian duration relations. Institute of Linguistics, University of Stockholm, PERILUS 13: 57–60.

Krull, D. 1992. Temporal and tonal correlates to quantity in Estonian. Institute of Linguistics, University of Stockholm, PERILUS 15: 17–36.

Lehiste, I. 1960. Segmental and syllabic quantity in Estonian. Indiana University, American Studies in Uralic Linguistics 1: 21–82. Bloomington: Indiana University Press.

Lehiste, I. 1965. The function of quantity in Finnish and Estonian. Language 41 (3): 447–456.

Lehiste, I. 1972. Temporal compensation in a quantity language. Proceedings of the Seventh International Congress of Phonetic Sciences: 929-939. Montreal, 22–28 August 1971. The Hague, Paris: Mouton.

Lehiste, I. 1975. Experiments with synthetic speech concerning quantity in Estonian. Congressus Tertius Internationalis Fenno-Ugristarum, Tallinnae habitus 17.–23.08.1970. Vol. 1: 254–269. Tallinn: Valgus.

Lehiste, I. 1985. An Estonian word game and the phonematic status of long vowels. Linguistic Inquiry 16 (3): 490–492.

Lehiste, I. 1988. Current debates concerning Estonian quantity. FUSAC '88. Proceedings of the Sixth Annual Meeting of the Fenno-Ugric Studies Association of Canada: 77–86. Lanham, New York, London: University Press of America.

Lehiste, I. 1990. Moraic analysis of Finnish and Estonian. Congressus Septimus Internationalis Fenno-Ugristarum, Sessiones sectionum 3B: 286–291. Debrecen.

Lehiste, I. 1995. Meetrika foneetikast. Keel ja Kirjandus 8: 525–533, 9: 596–603.
Lehiste, I., R. A. Fox. 1992. Perception of prominence by Estonian and English listeners. Language and Speech 35 (4): 419–434.
Liiv, G. 1961. Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid. Keel ja Kirjandus 7: 412–424, 8: 480–490.
Maddieson, I. 1993. Splitting the mora. UCLA Working Papers in Phonetics 83: 9–18. Los Angeles.
Posti, L. 1950. On quantity in Estonian. Suomalais-ugrilaisen Seuran Aikakauskirja LIV (2): 1–14. Helsinki.
Prince, A. S. 1980. A metrical theory for Estonian quantity. Linguistic Inquiry 11 (3): 511–562.
Ravila, P. 1962. Quantity and phonemic analysis. Proceedings of the Fourth International Congress of Phonetic Sciences: 490–493. Helsinki 1961. The Hague: Mouton.
Remmel, M. 1975. The phonetic scope of Estonian: Some specifications. Preprint KKI-5. Tallinn.
Remmel, M., A. Eek. 1971. The laterals: An acoustical study. The Ninth Acoustics Conference, Physiological Acoustics and Psychoacoustics: 138–141. Bratislava.
Shockey, L., R. Gregorski, I. Lehiste. 1971. Word unit temporal compensation. The Ohio State University, Working Papers in Linguistics 9: 146–165.
Tauli, V. 1973. Quantity and accent in Estonian. Commentationes Fenno-Ugricae. Suomalais-ugrilaisen Seuran Toimituksia 150: 390–403. Helsinki: Suomalais-ugrilainen Seura.
Trubetzkoy, N. S. 1939. Grundzüge der Phonologie. Prague.
Vihman, E. 1974. Estonian quantity re-viewed. Foundations of Language 11: 415–432.
Viitso, T.-R. 1963. Üks lingvistiline ahelprobleem. Nonaginta: 276–283. Tallinn: Emakeele Selts.
Viitso, T.-R. 1981. Läänemeresoome fonoloogia küsimusi. Tallinn: ENSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.
Viitso, T.-R. 1982. = T.-R. Vijtso, Moroschitayushtshij li yazyk estonskij? Soviet Finno-Ugric Studies 18 (1): 8–20.
Viitso, T.-R. 1986. Alternatiivsed fonoloogilised tõlgendused. Keel ja Kirjandus 7: 412–416.
Wiik, K. 1985. Regelsynthese zur Lautquantität im Estnischen. Ostseefinnische Untersuchungen. Ergebnisse eines Finnisch-Sowjetischen Symposions (Ed. H. Leskinen): 129–137. Helsinki: Suomalaisen Kirjallisuuden Seura.
Wiik, K. 1995. The Baltic Sea prosodic area revisited. The Fenno-Baltic Cultural Area (Ed. S. Suhonen). Castrenianumin toimitteita 49: 75–92. Helsinki.

Notes

[1] Patrik Bye's paper presented at the same Symposium has inspired us in completing the final version of the Conclusions.

[2] /sata/1 'hundred', nom.sg.; /palat/2 'pieces', nom.pl.; /palatt/3 'hospital ward', nom.sg.; /saatat/4 'send', 2.sg.pres.; /laulut/5 'songs', nom.pl.; /tappat/6 'kill', 2.sg.pres.; /kartat/7 'fear', 2.sg.pres.; /saatte/8 'get', 2.pl.pres.; /võitte/9 'you can', 2.pl.pres.; /ruuttul/10 'diamond', adess.sg.; /kaartte/11 'windrows', gen.pl.; /saatett/12 'sending, broadcast', part.sg.; /laulust/13 'song', elat.sg.; /tappaltt/14 'from Tapa', ablat.sg.; /kartaks/15 'he would fear', 3.sg.pres.condit.; /ruuttutt/16 'diamond', part.sg.; /saa:tus/17 'product', nom.sg.; /lau:lu/18 'song', part.sg.; /tap:pa/19 'to kill', da-infinitive; /kor:tav/20 'repeating', pres.participle, nom.sg.; /saat:te/21 'sending, broadcast', gen.sg.; /võit:ta/22 'to win', da-infinitive; /kaart:te/23 'cards', part.pl.; /pois:se/24 'boys', part.pl.; /loo:tta/25 'story', abess.sg.; /või:tta/26 'butter', abess.sg.; /poi:sse/27 'buoy', illat.sg.; /hoo:kne/28 'in full swing', nom.sg; /saa:tutt/29 'that which has been got', part.sg.; /saat:teks/30 'sending, broadcast', translat.sg.

Mõned lihtsad tajukatsed eesti sõnaprosoodia alalt: kas jala struktuur või segmentaalne kvantiteet?

Arvo Eek ja Einar Meister

Käesolev artikkel koosneb kolmest põhiosast. Esimeses osas antakse ülevaade eesti sõnaprosoodia kohta teadaolevast andmestikust. Andmete analüüs kergitab uusi küsimusi ja suunab tähelepanu mõnele vanale, kuid seni rahuldava lahenduseta probleemile. Mõnda nendest küsimustest testitakse tajukatsetega artikli teises põhiosas.

Meil õnnestus tõestada, et kuulajad on suutelised segmentaalsel ja süllaabilisel tasandil määratlema ainult kaht (lühikest ja pikka), mitte kolme distinktiivset väldet. Traditsioonilise kolme silbivältena tuntud nähtust, mida siin aktsentidena käsitleme (A1, A2, A3), on võimalik tajuda ainult siis, kui kuulajale on kättesaadavaks tehtud ka teise silbi vokaalis peituv teave (katse 1, joonis 1). Silp ei ole eesti keeles väikseim distinktiivne prosoodiaüksus. Selleks osutub kahesilbiline aktsendijalg. Aktsentide esinemus sõltub sõna rõhumallist. Iga rõhujalg, olgu pea- või kaasrõhuline, on mõne kindla aktsendina identifitseeritav. Aktsentide äratundmiseks oluline rõhujala osa koosneb rõhusilbi riimist (A2 ja A3 korral minimaalselt selle lõpuosast) ning rõhutu silbi tuumast (katse 2). Seda rõhujala osa nimetatakse siin aktsendijalaks. Rõhujala kolmas silp (kui see juhtub olema), silbialgused ja rõhutu silbi lõpukonsonant jäävad aktsentide määratlemisel inertseteks. Aktsentide identifitseerimiseks olulisima informatsiooni koondumine mõlemale poole silbipiiri tekitab tajumulje A3 harjalisusest (energeetilisest tasakaalustamatusest) ja A2 tasakaalustatusest. A3 jalg on kas vokaal- või konsonantharjaline. Sellise tajumulje tekitamisest võtavad kestussuhete kõrval osa ka põhitoon ja intensiivsus (katse 3). Tasakaalustamatu A3 jala harjalisuse efekt võimendub lühenenud, nõrgenenud ja kvalitatiivselt redutseerunud teise silbi vokaali taustal. Ka A2 kahesilbilise jala tasakaalustatus tuleb esile pikenenud ja kvalitatiivselt redutseerimata teise silbi vokaali taustal.

Rõhujala piires on täheldatav kaks temporaalse regulatsiooni mehhanismi. Kogu rõhujalaga (mis saab olla maksimaalselt kolmesilbiline) seotud isokrooniline suunitlus väljendub selles, et häälikute lisamine jalale lühendab jala koostissegmentide kestusi. See mehhanism tekitab üldtausta ega ole aktsentide määratlemisel oluline. Rõhujalale allutatud madalama tasandi temporaalse regulatsiooni mehhanism iseloomustab aktsendijala isokrooniat. Rõhust tingitud isokrooniline suundumus mõjutab segmentide absoluutseid kestusi. Aktsentidest põhjustatud silpidevaheline temporaalne regulatsioon väljendab mõlemal pool silbipiiri olevate segmentide kestuste aktsenditüübile iseloomulikku suhtestatusse viimist (pöördvõrdelised kestussuhted), nii et seejuures aktsendijalad jäävad aktsenditüübist sõltumata isokroonseteks (tabelid 4, 5, 6).

Töö lõpuosas esitatakse põhimõistete ja nendest tulenevate piirangute raamistik eesti keele sõnaprosoodia hierarhiliseks käsitluseks (joonis 2). Tuginedes sellest kontseptuaalsest raamistikust tulenevale moorakäsitlusele, esitame joonisel 3 kokkuvõttena eesti aktsentide põhilised jalatüübid.

Estonian Prosody: Papers from a Symposium

EDITED BY ILSE LEHISTE AND JAAN ROSS

Proceedings of the International Symposium on
Estonian Prosody,
Tallinn, Estonia, October 29-30, 1996

Institute of Estonian Language TALLINN

Publication of this book has been supported by the Open Estonia Foundation and the Eduard Treu Memorial Fund.

Cover and title by Heiki Looman
Layout by Jüri Helekivi

ISBN 9985–851–47–1

Estonian Prosody: Papers from a Symposium

EDITED BY ILSE LEHISTE AND JAAN ROSS