# THE INTERRELATIONSHIP BETWEEN PHONOLOGICAL AND PHONETIC SOUND CHANGES: A GREAT RHYTHM SHIFT OF OLD ESTONIAN

Arvo Eek, Toomas Help

Institute of Language and Literature,
Estonian Academy of Sciences,
Tallinn 200106, USSR

## 1. PURELY PHONETIC SOUND CHANGES

In the literature on historical linguistics, phonological and phonetic sound changes are often contrasted. A *phonological change* alters the phoneme composition of certain speech flow appearances, a *phonetic change* alters the way in which different expiratory-phonatory-articulatory gestures and their acoustic-auditory correlates are presented in certain speech flow appearances. For instance, in Old Litomyšl Czech, there was a sound change **p, > t* , manifested e.g. in **p,i:vo > ti:vo* 'beer'. It was a phonological change in the sense that the phoneme /p,/ was exchanged for the phoneme /t/ by replacing (in terms of Andersen 1973) the distinctive feature 'heightened low consonant tonality' of /p,/, by the feature 'high consonant tonality' of /t/. It was a phonetic change in the sense that the actual labial articulation of [p,] characterized (in terms of Ladefoged 1980) by certain values of the articulatory parameters like 'lip height', 'lip width', and 'lip protrusion' and its corresponding 'heightened low tonality' in acoustics-audition were exchanged for the actual dental articulation of [t] characterized by certain values of the articulatory parameters like '(tongue) tip raising' and '(tongue) tip advancing' and its corresponding 'high tonality' in acoustics-audition.

In historical linguistics all relevant sound changes are traditionally reduced to phonological changes. The prevalent strategy of historical linguists has long been to conceive the main course of a sound history as a *chain* of phonological changes and to view phonetic changes as mere detailed specifications of the stated phonological changes. Any single phonetic change, relevant to the main course of the sound history, has been viewed as obligatorily subordinated to a concrete phonological change. The traditional strict parallelism between phonological and phonetic changes is revealed in the theoretical framework of Andersen 1973, the original source of the above Old Litomyšl Czech example. Here the speciality of phonetic changes is emphasized to the extent that 'abductive' phonological and 'deductive' phonetic changes are dealt with as two different categories (about the equation of 'abduction' to phonology and 'deduction' to phonetics cf. p. 774). The framework articulates that a concrete historical sound change is always a complex of a phonological change and a phonetic change, i.e. phonological changes are understood as abductive innovations in the mental coding of sounds and phonetic changes as subsequent deductive innovations in their physico-physiological manifestation.

The traditional focus on phonological changes only rather than on phonetic changes as well, seems to be due to two methodological aspects of historical linguistics. On the one hand, most methods of historical sound reconstruction imply the view that sounds are *discrete* units: they rely on comparisons and differentiation between lexicon units (words, morphemes). On the other hand, only phonological changes are easy to be conceived as changes of discrete units (phonemes, distinctive features), whereas phonetic changes are better to be conceived as *continuous* changes of parameter values (determiners of the actual articulatory gestures and their acoustic-auditory correlates).

We argue that some phonetic changes, relevant to the main course of a sound history, have no concrete phonological changes as one-to-one support. Let us call them *purely phonetic changes*. We conceive the main course of a sound history as a chain of changes some of which allow both phonological and phonetic specifications but some phonetic specifications only. Consequently, we do not reduce all relevant sound history to its phonological changes. The proposed equal focus on phonological and phonetic changes makes the recognition of historical sound changes more sophisticated. In order to reconstruct purely phonetic changes, the discrete philosophy of the method has to be adapted to the continuous nature of changes. However, these methodological complications do not hinder us in attributing the phonetic changes the essential ontological value. Instead, we view them as an inevitable forfeit to be paid for the proposed theoretical adjustment.

The essence of purely phonetic sound changes is exemplified by introducing a Great Rhythm Shift of Old Estonian. However, before going to the change itself, we outline a theoretical framework in which the



essentials of speech rhythm are treated in purely phonetic terms so that two optional rhythm strategies - striving vs. switching foot control - are extracted.

## 2. STRIVING VS. SWITCHING CONTROL OF FOOT

### 2.1. *Phonological vs. phonetic perspective on language*

According to Ladefoged 1984, the role of phonemes for individual speakers and listeners has been grossly exaggerated in the tradition of language research. Phonemes, distinctive features and other units of this size are empirically valid devices of speech description only if language is viewed as a *social* norm of a community. They are too abstract devices while describing language as a *psychological* act or state of an individual, e.g. while describing the actual realization or the mental coding of concrete speech samples. To draw on his original parallel, phonemes are like moral prescriptions or economical laws: they are manifested in the behaviour of a human group rather than a single member of this group.

In concrete linguistic analysis, the social rather than psychological nature of phonemes is revealed best by the characteristics of a boundary between two subsequent segmental phonemes. In the psychological perspective, the boundary is a fiction: if to observe one single speech flow sample alone, no invariant cues of a boundary between phoneme-size segments are available (cf. Hammarberg 1976). The boundaries become a reality in the social perspective. Here, the observed speech flow sample is confronted with a set of potential resembling samples and only its recurrent details are extracted as relevant features. Similarly, a suprasegmental phoneme is evident in a set of possible speech flow samples rather than in a single observed sample.

The relationship between phonology and phonetics leads to a crucial revaluation. Phonology and phonetics do not deal with different empirical data, they are not complementary portions of a unitary description system. In particular, the mental coding of speech is reflected not only by phonology and the articulatory-acoustic-auditory realization of speech is reflected not only by phonetics. Rather, phonology and phonetics deal with the same empirical data but represent two essentially incompatible strategies that are proper for different purposes.

*Phonology* describes both the mental coding and the articulatory-acoustic-auditory realization of speech flow in the *functional* perspective. Here, speech flow is viewed as a combination of details that are selected from the total set of observation results as these features that motivate the existence of the particular phenomenon 'speech flow'. Speech flow exists in order to convey linguistic meaning, i.e. to indicate which linguistic units (morphemes, lexical words etc.) are actually used one after another in social communication.[1] Consequently, the functional perspective entails the view that speech flow is a string of phonemes, i.e. abstract meaning-differentiating capacities that are evident when speech flow equivalents of different linguistic units contrast with each other.

*Phonetics* describes both the mental coding and the articulatory-acoustic-auditory realization of speech flow in the *formal* perspective. Here, speech flow is viewed in all its details available in the observation of a speech sample as a psycho-physiological and acoustic act. According to Ladefoged 1984, phonemes, distinctive features, and other units of this size do not belong to such empirical realities. Instead, the formal perspective entails the view that speech flow is a continuously produced energy wave in the sense that it does not split into segmented units of phoneme size. In this context we may follow Plomp 1984 in identifying the acoustic appearance of speech flow as an air flow supported by continuous activity of respiratory mechanism and radiated from a human being as a wide-band signal that is modulated continuously in time by manipulating vocal cords (fundamental frequency), by narrowing and widening the vocal tract (temporal intensity envelope), and by modifying the vocal tract cavities (frequency spectrum), and received by the peripheral hearing apparatus. We may change the perspective and claim that speech flow appears as a continuous *speech energy* that passes subsequently through three media. First, it is produced in the physiological medium as the mentioned four-fold fluctuations in speech organs (speaker's respiratory, phonatory, and articulatory activities); second, it is transmitted in the physical medium as a modulated wide-band signal (acoustics); and third, it is received in the physiological medium as a fluctuation in the peripheral hearing apparatus (listener's auditory activation).[2]

Speech rhythm is a phenomenon that is directly manifested in speech flow: speech flow is actually a continuous alternation of the minimal and maximal levels of speech energy. We have posed phonetics and phonology as two mutually exclusive research strategies. In this context, we try to fix the essentials of speech rhythm in purely phonetic, energetical terms without invoking on phonological consideration.

### 2.2. *Stress, foot, syllable and demisyllable: a model of speech rhythm*

Explaining historical sound changes and describing typological differences in contemporary languages we proceed on the following crude model of speech rhythm.

*Stress* is the total energy amount spent by the speaker's complex expiratory, phonatory and articulatory activities (gestures) while producing a stretch of speech flow (for a review of literature on stress production and acoustics: Eek 1982; MacNeilage 1972) rather than some special reinforcing energy added to a certain independently defined unit of speech flow. Stress itself is intrinsically segmented into units rather than appears something like an energetical increment of stressed syllables relative to unstressed syllables.

*Foot* is the minimal integral unit of stress. Foot organizes speech energy into a stress impulse; its *general* shape, i.e. a hypothetical temporal energy envelope, is physiologically determined by the tension/relaxation phases, inevitably needed in the activities of speech organs, and its *detailed* shape is specified by language-specific commands on speech organs (Fig. 1). Foot cannot be understood as a chest pulse, a unit of laryngeal fluctuation or a motor unit like the articulatory syllable of Чистович, Кожевников et al. 1965, it is temporally organized amount of speech energy produced by all the activities.

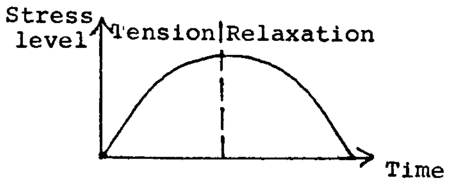


Figure 1. A model of foot.

In particular, it is foot that is best to be conceived as the domain of *accentuation*, i.e. the purposeful variation in the amount of stress. Psycholinguistic experiments (e.g. Terken 1984, cf. also Scott 1982) indicate that the amount of stress displayed in speech production is in a direct correlation with the importance of the signal in the interpretation of the corresponding message. A speech signal is maximally accentuated, if it alone serves as the basis for computing the meaning of the corresponding message; a speech signal is minimally accentuated, if the preceding linguistic context and the extralinguistic state of affairs rather than the speech signal itself serve as the basis for computing the meaning.[3] At the maximum end of the accentuation scale, there are *emphasized feet* that show an integral portion of stress of the largest size. In this case the needed amount of stress is warranted by special reinforcing activities including registerable pulmonic activities (cf. Ohala, Riordan, Kawasaki 1979). At the minimum end of the accentuation scale, there are *tonic feet*. In this case the needed amount of stress is warranted by keeping to the speech organs activities that are necessary and sufficient for the speech signal transmission. Between these two extremes there remain simply *accentuated feet*.

The intrinsic mode of stress appearance is a foot. In this physiologically predetermined limit the detailed shape of energy distribution is controlled by both universal and language-specific segmental and suprasegmental commands.

Speech is decomposed into continuous alternations of the narrowing (i.e. consonantal) and the widening (i.e. vocalic) gestures in the vocal tract. As the first approximation we may suppose that the detailed time-ordered segmental specifications for the basic CV- and -VC *demisyllable-gestures* are controlled by *demisyllabic commands* (cf. Fujimura 1983). Demisyllabic command scans all muscle channels exploited in generation of the respective minimal articulatory sequence and turns on simultaneously these channels which activity is not contradicted to the concrete state of the movement (cf. coarticulatory phenomena e.g. in the sequence of *plu-*: during the articulation of *p* lips are already rounded, immediately before *p* release the tongue tip is raised to the alveolar region and the postdorsum is moved towards the velum). Demisyllabic commands regulate articulatory movements in the speech flow. Demisyllable as a unitary articulatory sequence generates an indispensable (inherent) amount of speech energy that is sufficient for the complex gesture; its duration is determined by inherent durations of the combined movements. Consequently, demisyllables themselves do not specify the shape of a temporal energy envelope (i.e. the shape of a foot), they simply divide a foot into minimal inherent energy blocks.

Detailed modifications of a stress impulse may be controlled by two kinds of hypothetical energetic commands. These commands - switching and striving - determine the turning point between the tension and relaxation phases (cf. Fig. 1) of the temporal energy envelope.

If the *switching*-type foot control is used, there are usually two switching commands in a foot. The first of them turns on a stress impulse and predetermines the quickest way to the maximal energy level (thus controlling mainly the tract-widening segment, i.e. the vowel quality), but does not define the temporal characteristics of the maximal energy level itself; duration of the latter follows mainly from the inherent temporal properties of the corresponding demisyllable(s). Physiologically conditioned relaxation is turned on by the second switching command, thus showing the location of the turning point of a foot and simultaneously guaranteeing relatively exact sound quality. Thus, any switching command entails a rapid rise to an energy level. The switching-type foot control predicts a relatively tight coarticulation before the initial point of the maximal energy level of a foot and relatively loose connections between underlying segments after that point. As a rule, a foot tends to be a isobaric (in energetic terms) or isochronic unit (in terms of timing). In the case of



switching the isobaric/isochronic tendency to an average strength or length value of all feet would be revealed relatively weakly.

If the *striving*-type foot control is used, there is one striving command in a foot. A striving command turns on a stress impulse smoothly and its main force is directed to the maximal energy level of a foot. The relaxation, i.e. the decline towards the minimal energy level, begins automatically after the final point of the tension phase is reached. In this case one would expect a relatively loose connections between underlying segments at the beginning of a foot (e.g. a diphthongization-like vocalic gesture may appear there) and close contact immediately before the exactly defined final point. A quality reduction of the underlying vocalic gesture is predictable in energetically uncontrolled relaxation phase. As the controllable stretch of a foot is essentially longer than in the switching case, a stronger isochronic tendency would appear.

If we want to delimit intra-foot segments that are influenced by energetic commands more properly we must define a syllable-size unit. Articulatorily, *syllable* is a relatively homogeneous sequence of demisyllables delimited by opening gestures of the vocal tract. This definition relies on two findings. On the one hand, the essential energetic commands - striving and switching - are inclined to apply to these demisyllabic commands at which a transition to the vocal tract opening gesture begins. On the other hand, these demisyllabic commands that are directed to a closing gesture are usually under the influence of preceding switching or striving commands. This regularity presents a crucial support to speech perception. Namely, in the corresponding acoustic sound wave used in the transmission of the message from speaker to listener, syllables are designated by the temporal amplitude envelope, i.e. the beginnings of syllables are definitely marked by transitions to the intensity rise of the sound wave. Thus, in acoustics a syllable is the foot stretch containing of one easily identifiable (cf. Mermelstein 1975) temporal amplitude envelope. The complex relationship between articulation and the produced sound wave acts as a temporally selective energy filter. For instance, the acoustic onset of a syllable does not immediately coincide with the relative time-point marked by the onset of the speech organs activity, i.e. by the onset of a foot (for the literature on perceptual centers see e.g. Fowler 1979, Marcus 1981, Howell 1984). However, on the basis of suggested close physiological interrelationship between speech production and perception (cf. the motor theory of speech perception of Liberman et al. 1962, the auditory-motor theory of speech production of Ladefoged et al. 1972; cf. also the observed parallels of syllable production and perception by Tuller and Fowler 1980) the information about an articulatory syllables may still be relevant in speech perception and vice versa.

Within the accentuation range that is determined by the functional characteristics of an utterance, the speech flow is produced by stress segments represented by foot-size alternations of the minimal-maximal-minimal physiological energy levels, i.e. by a sequence of temporal energy envelopes. In acoustics and perception, there are clearly identifiable intra-foot units characterized by separate temporal intensity envelopes we have defined as syllables. A *foot* may consist of *one*, *two* or *three syllables*. However, some accentuated or emphasized stress impulses may be conceived as displaying more than three syllables. In such cases we suppose that the accentuated or emphasized foot is followed by one or more intervening weak tonic feet (cf. Lea 1974: 41 for the phonetic justification of the 'one-two-three-syllables' principle).

### 2.3. *Types of rhythm organization*

Languages differ as regards the method by which energetic stress commands apply to their feet. Generally, the most "natural" foot types seem to be those in which every single syllable is affected by one command only. There are two "natural" types of languages that exploit this principle as a rule, i.e. display foot control systems subordinated to syllable-switching or syllable-striving.

1) *Syllable-switching* foot control is characteristic of languages with prototypical polysyllabic feet (e.g. in Finnish and Italian); in this type of languages monosyllabic feet are rarely used.

Syllable-switching languages give much attention to the beginnings of syllables. As a result of that the target of the following vowel is clearly defined, there are no perceivable diphthongization of short vowels; in the regulation of stress degrees also foot-initial consonant may take a part (cf. e.g. data for Tamil: Balasubramanian 1979). In foot-final syllables, syllable-switching languages have an open set of vowels to choose from (cf. e.g. the vowel harmony in Finnish) and do not, regularly, show reduced vowels, or more correctly, do show less reduced vowels than e.g. foot-striving languages (for Finnish: Wiik 1965, for Italian: Bertinetto 1981). Syllable-switching languages may have an opposition between short and long segmentals. This language type displays a relatively weak foot-level and syllable-level isochrony (for Italian: Vayra, Avesani, Fowler 1984, for Finnish: Lehtonen 1970 and for Tamil: Balasubramanian 1979). Here the inherent endeavour to foot isochrony may be strongly reduced, because the duration of a foot is determined roughly by their inherent temporal properties.

2) *Syllable-striving* foot control is characteristic of languages with prototypical monosyllabic feet (e.g. Vietnamese).

In syllable-striving languages, polysyllabic words cannot be analyzed into a continuous alternation of stressed and unstressed syllables (cf. Касевич 1983) due to foot monosyllabicity: here relaxation occupies a final part of a syllable rather than a separate syllable. Syllable-striving languages are usually tone languages, since the control over the vowel matter and the whole maximal energy segment develops conditions for tonogenesis.

The principle according to which every single syllable is affected by one energetic command may be violated in two additional, less "natural" types of languages. Here the "natural" foot types are accompanied by less "natural" types. In languages that display foot control systems subordinated to foot-striving, there are feet in which some syllables are left without any energetical control altogether. In languages that display demisyllable-switching, there are feet in which one syllable is affected by two separate energetic commands.

3) *Foot-striving* control is characteristic of languages with mono- and polysyllabic feet (e.g. English, German, Russian).

All that has been said about syllable-striving languages may also be valid for foot-striving languages and vice versa (the only exception being phenomena related to non-foot-initial syllables of foot-striving languages).

Striving command does not pay special attention to the beginning part of a syllable (a reason of possible diphthongization), it is directed to the place of the maximal energy point either on a vowel, a consonant matter or on some consonant in a consonant cluster (a reason for allowing relaxed syllable affixes, cf. Fujimura, Lovins 1978). At the same time, eligibility of the maximal energy point for different syllable segments in different words usually changes the character of short/long segmental opposition known in syllable-switching languages for a complementary distribution type of short/long opposition between vowels and consonants, thus guaranteeing fairly convenient conditions for foot isochrony (a strong negative correlation between neighbouring vowel and consonant, sometimes defined as loose vs. close contact). Foot-striving languages have a strong tendency towards foot isochrony. Data on prominent intra-foot temporal compression are available from a number of foot-striving languages (for English: Klatt 1973, Fowler 1981; for Swedish: Lindblom, Rapp 1973; for Dutch: Nooteboom 1972). A characteristic feature of foot-striving languages is the foot-final vowel reduction (for English: Wiik 1965, for Russian: Бондарко et al. 1966, for Swedish: Lindblom 1963). As for Russian, experimental data corroborate the model-predicted uncontrollability of the foot-final syllable quality. Here the full complexity of reduction cannot be established through the study of only the vowels themselves; the analysis of stressed and unstressed syllables has revealed that with the intensification of reduction the degree of coarticulation changes (Бондарко et al. 1966). Experiments with varied speech tempo have shown that at fast speaking rates the quality of the stressed vowel is not susceptible to qualitative reduction. The tendency to reduce the formant frequencies only shows up in unstressed vowels, and it does so even when the unstressed vowel has the same duration as its fast-rate stressed counterpart (Зиндер 1964). It is in accordance with the viewpoint that duration and articulatory effort may be controlled separately.

4) *Demisyllable-switching* foot control may emerge in languages with mono- and polysyllabic feet (e.g. Japanese; we consider e.g. the word *Sapporo* as consisting of a monosyllabic foot *sap-* and a polysyllabic foot *-poro*)

In the case of monosyllabic feet the turning point between tension and relaxation phases can be controlled in two principally different ways, i.e. by striving or switching commands. In Japanese, the striving possibility is refuted by the peculiarities of word rhythm (cf. Homma 1981). However, the domain of a syllable-switching command is the syllable beginning only and it cannot apply to the end of a monosyllabic foot. In this context, switching has to be shown on demisyllables both in and outside syllable beginnings, converting syllable-switching into demisyllable-switching. In short, a Japanese monosyllabic foot has to be controlled by two demisyllable-switching commands. As much as the temporal organization of a demisyllable is dependent upon inherent timing properties of underlying segments, we may expect that demisyllables represent temporally more or less equal units. In Japanese, where a demisyllable may be interpreted as a mora-size unit, Sawashima et al. 1982 have reported that the relative timing of articulatory and vocal pitch control is organized so as to compensate for timing variations in the internal mechanisms and maintain constant temporal relations in the acoustic output. The equality of the demisyllabic units is supported by the perceptual data (cf. Fujisaki, Horiguchi 1979). Modern Japanese, a language with the demisyllable-switching foot control, has neither isochronic feet (cf. the parallel existence of feet of two and three moras) nor syllables (cf. data in Homma 1981).

5) *Compound* foot control may emerge, inter alia, in languages with mono- and polysyllabic feet (e.g. Estonian).

We have presented the above four strategies as mutually exclusive options a language may follow. Note that our argumentation relied on the assumption that the chosen strategy is a mere inevitable answer to a physiological constraint, the alternation of tension and relaxation in feet, not directly used in meaning differentiation.





However, the strict boundary between the types may vanish in languages with contrastive meaning-differentiating accents, in case of which a special attention is paid to the energetic behaviour itself (cf. the psycholinguistic parallel of sound repetitions across morpheme boundaries that reveal the same twofold treatment: Menn, MacWhinney 1984).

In Estonian, in a quantity and accent language, syllable-switching and foot-striving control strategies together constitute an integral whole.

On the one hand, some essential features of Standard Estonian refer to the syllable-switching foot control: a) foot-initial short vowels are not diphthongized; b) non-foot-initial vowels do not show quality reduction; c) all 9 vowels and 17 consonants may occur as short or long phonemes (traditionally treated in terms of three quantity degrees, cf. Ariste 1938); a short/long opposition does not display any rules of complementary distribution (for the distribution of short and long phonemes see Eek 1986).

On the other hand, there are some substantial characteristics of foot-striving in Standard Estonian: a) all feet occur in either an even or in a sharp accent (see Table 1; for productional and acoustic data on accents see Eek 1986); b) a strong temporal compression is supported by the fact that mono- and disyllabic feet, irrespective of the accent type and segmental duration, differ durationally less than do intra-feet segments; among all segments that constitute a foot there exist a significant temporal relationship (cf. Lehiste 1972, Eek 1974); the foot tends to be as an isochronic unit of a temporal program (cf. Eek, Remmel 1974); c) the occurrence of vowels in non-foot-initial position is restricted, etc.

Note: in the discussion below, we use the following designations: $\rightarrow$ - a demisyllabic command, $\uparrow$ - a switching command, $\curvearrowright$ - a striving command.

Table 1
The appearance of the opposition 'even vs. sharp accent' in modern Standard Estonian

| | even accent ´ (foot consists of 2 - 3 syllables) | sharp accent ` (foot consists of 1-3 syllables) |
|---|---|---|
| polysyllabic foot with a long first syllable | *k´auna* 'pod. GENITIVE' | *k`auna* 'pod. PARTITIVE' |
| polysyllabic foot with a short first syllable | *k´ana* 'hen' | Ø |
| monosyllabic foot (obligatorily long syllables) | Ø | *k`aun* 'pod' |

2.4. *Stress-timing vs. syllable-timing*

Pike 1946 and Abercrombie 1967 are among the principal works that introduced the distinction between stress-timing and syllable-timing, two mutually exclusive, essentially different types of speech rhythm that a language may show. In a stress-timed language like English, stress beats were assumed to recur at approximately equal time intervals, in a syllable-timed language like French, syllables were assumed to recur at approximately equal time intervals. The two categories are viewed as mutually exclusive on the assumption that intervals between subsequent stress beats may be filled invariantly with a varying number of syllables.

The distinction between stress-timed and syllable-timed languages reflects undoubtedly some optional fundamental qualities of speech rhythm: note that it has been made use of in a large number of papers dealing with very different languages (cf. e.g. Lehiste 1977 and Dauer 1983 for a bibliography). However, the whole issue has been labelled as a linguistic controversy (Roach 1982). First, phoneticians do not agree with the view on speech production that the distinction implies: all syllables cannot be associated with separate chest pulses and stressed syllables cannot be extracted as special reinforced chest pulses (Ladefoged 1968). Second experimental psycholinguists deny the reality of any measurable timing difference between the rhythm types even in the prototypical opposition of English and French (Scott, Isard, Boysson-Bardies 1985). Third, both naive and expert listeners in experimental situations and comparative linguists in their theoretical treatments often disagree in attaching a particular language to either of the categories or, instead, claim that it belongs to neither (cf. Miller 1984).

To abandon these contradictions, we propose the model of speech rhythm in which 'stress' and 'foot' are considered as basic notions rather than 'stress' and 'syllable'. First, as for special short-term pulmonic activities in speech production, a foot rather than a syllable could be regarded the minimal integral domain to which they may apply (emphasized foot) but by no means need apply (tonic foot). Second, as for the basic non-timing nature of the distinction, the actual temporal rhythm pattern of a language could be viewed as deriving from two essentially non-temporal appearances of feet. Whether a language is stress-timed or not, depends on the interrelationship between feet and stress beats, accurately, on the frequency in which feet appear in continuous speech flow in the accentuated (reinforced) form displaying thus stress beats. Whether a language is syllable-timed or not, depends on the interrelationship between feet and syllables, accurately, on the manner in which syllables associate with the internal structure of feet. Third, as for the actual fuzzy boundary between the rhythm types in cross-linguistic research, the foot perspective entails a much more complicated picture of factor interplay than a strict twofold opposition between stress-timing and syllable-timing languages.

2.4.1. *Stress-timing: general motor rhythm vs. temporal speech rhythm*

Allen 1975 emphasizes the distinction between *general motor rhythm*, i.e. a pattern of sequence, and *temporal rhythm*, i.e. a pattern of temporal sequence. We have already introduced speech rhythm in the sense of a general motor rhythm, a successive sequence of feet. However, speech flow is subject to a universal bias towards a temporal rhythm as well. The portion of accentuated feet, displaying the highest energy levels, are conceived as *stress beats*, i.e. speech energy concentration sites in a continuum of speech flow that usually display more or less equal energy distribution. Like other sequences of rapid movements in human behaviour (e.g. finger tapping), subsequent stress beats tend to recur often regular intervals according to the properties of a universal physiological temporal rhythm pattern (cf. Allen 1975). In particular, stress beats are inclined to cluster around a mean interval of 0.4 to 0.5 s with an overall range of interval durations limited to between 0.2 to 1 s.

Languages are apparently *stress-timed* if they show clearly the temporal pattern of stress beats superimposed on speech flow by the physiological temporal rhythm, otherwise they are *non-stress-timed* displaying rather an intrinsic timing of their feet. In other words, the distinction 'stress-timing' vs. 'non-stress-timing' reflects properties of the real inter-foot timing phenomenon that is revealed between feet in continuous speech flow.

In all languages, there are obligatorily some feet that show accentuation to the extent that exceeds the so-called *stress beat threshold*. A language shows stress-timing, if such beats in continuous speech flow are *frequent* enough to converge into a pattern of temporal rhythm, and does not show stress-timing if they are too *rare* for that, cf. Fig. 2.

In typological research, it is often easy to decide whether a particular language is stress-timing or not. On the one hand, English (Dauer 1983) and Polish (Biedrzycki 1980) are obvious stress-timing languages because they display salient stress beats recurring frequently around 0.5 s (as a rule, indicating functionally that a new lexical entry is present in message) and are thus subject to a clear temporal rhythm. However, it does not mean that the so-called syllable-timed languages, on the basis of mean interbeat intervals, cannot belong to the same group with English (cf. e.g. Spanish: Navarro 1932). If lexical stress is saliently marked and lexical entries are not too long (not exceeding 4-5 syllables or 2 tonic feet) there is high probability to perceive such syllable-timing language as a stress-timing language.

On the other hand, French and Japanese (Dauer 1983) are obvious non-stress-timing languages because they display salient stress beats rarely after longer intervals than 0.5 s (as a rule, indicating functionally that especially important lexical entries of a phrase, clause, or sentence are present in message) and thus are not subject to a temporal rhythm. Rather long distance between stress beats in these languages cannot be accounted for in pure functional terms. If a language is characterized by non-prominent word stress and at the same time, accentuation does refer to every lexical entry, we can perceive a non-stress-timing phenomenon.

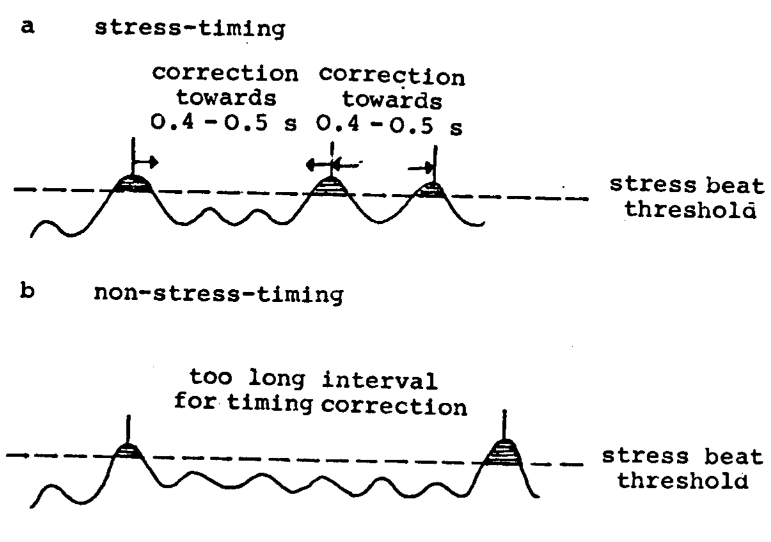


Fig. 2. Temporal rhythm superimposed (a) or not superimposed (b) on a sequence of feet.



We assume that it is because of the polarization between the non-stress-timing French and the stress-timing English that, in executing tasks of temporal rhythm manipulation, subjects who have French as their mother tongue show vagueness that is alien to subjects speaking English as first language (cf. data and discussion on the issue in Scott, Isard, Boysson-Bardies 1985). Nevertheless, we follow Dauer 1983 in claiming that whether a langue is stress-timed or non-stress-timed is a matter of degree rather than of mutual exclusion: many particular languages show neither obvious stress-timing nor obvious non-stress-timing but something between the extrema. For example, Finnish displays salient 'primary' stress beats (as a rule, indicating functionally that a new lexical entry is beginning in message) that occur too rarely in speech flow to converge into a pattern of temporal rhythm (in Finnish, an agglutinative language, the distance between the beginnings of subsequent lexical entries may be rather long). Consequently, Finnish is not subject to a clear temporal rhythm. However, a slight 'secondary' stress beat is provided by all other Finnish stress impulses as well so 'primary' and 'secondary' stress beats together are frequent enough to converge into the pattern of temporal rhythm. Accordingly, Finnish is still subject to a dim temporal rhythm (cf. the treatment of the Finnish problematics in O'Connor 1973).[4]

Table 2 summarizes our argumentation.

Table 2
The scale of stress-timing

| less stress-timing ← | | → more stress-timing |
|---|---|---|
| French<br>Japanese | Finnish | English<br>Italian |

2.4.2. *"Syllable-timing"*

In polysyllabic feet of many languages, the average duration of a non-final syllable is generally under 0.2 s. In this context, syllables cannot be fundamentally subject to the general temporal rhythm pattern like that of stress beats (Dauer 1983) and as it is supposed by the term "syllable-timing".

We suggest that the term "syllable-timing" stands simultaneously for two different properties of the internal structure of feet the main common feature of which is the fact that they are alien to English, the prototype of "non-syllable-timing" languages. On the one hand, a language is "syllable-timed" if it displays a switching foot control. On the other hand, a language is "syllable-timed" if it has always one syllable for one striving or switching command (Table 3).

Table 3
The combinations of "syllable-timing"

| | 1) switching rather than striving | 2) one switching/striving command for one syllable rather than one demisyllable or many syllables | |
|---|---|---|---|
| Finnish | + | + | ↑ more "syllable-timed" |
| Japanese | + | - | |
| Vietnamese | - | + | |
| English | - | - | ↓ less "syllable-timed" |

3. THE GREAT RHYTHM SHIFT: OLD ESTONIAN

3.1. *Conservative Finnish vs. innovative Estonian*

In a number of general works on language (e.g. Anttila 1972, Comrie 1981), the comparison of the two main Balto-Finnic languages, Finnish and Estonian, serves to illustrate the point that genetically close-related languages may differ remarkably in respect of their typological characteristics. In outline, Finnish has preserved the original fairly clear-cut agglutinating morphology but Estonian has exchanged it for a morphology that is much more strongly characterized by fusion. The morphological differences are accompanied by crucial differences in the sound architecture of the languages. Finnish has preserved firmly the original Balto-Finnic prototype of long polysyllabic words that consist of simple syllables of the structure CV or CVC, display an extensive vowel harmony, and begin at an accentuated foot of an invariable quality. On the contrary, Estonian has introduced many short mono- and disyllabic words, complex syllables like CCVCC, and word-level restrictions on vowel distribution that have abandoned the original vowel harmony altogether. Estonian accentuated foot is mobile (a word need not begin at an accentuated foot), and variable (it displays either the even or sharp version of the contrastive accent).

These essential differences between modern Finnish and modern Estonian originated with a row of phonological changes that, on the one hand, occurred in the history of Estonian during the first centuries of the second millenium A.D. (roughly, 1100 - 1500) but, on the other hand, were absent in the history of Finnish. This claim has a high degree of confidence as it is supported by historical and comparative linguistic evidence and early textual data on Estonian. We concentrate on some central component changes of the row, cf. Table 4.

The general pattern of the below presented phonological changes includes an essential

Table 4
Some crucial phonological changes of Old Estonian
in 1100-1500

(1) RISE OF CONTRASTIVE ACCENTS
(´ = EVEN ACCENT, ` = SHARP ACCENT)

| | | |
|---|---|---|
| **kaunan* 'chaff/pod.GENITIVE' | > | **k´aunan* |
| **kaunaan* 'chaff/pod.ILLATIVE' | > | **k`aunan* |
| Finnish *kaunan* | | Estonian *k´auna* |
| *kaunaan* | | *k`auna* |

(2) SHORTENING OF LONG VOWELS
IN NON-INITIAL SYLLABLES

| | | |
|---|---|---|
| **mustaa* 'black.PARTITIVE' | > | **musta* |
| Finnish *mustaa* | | Estonian *m`usta* |

(3) LOSS OF SHORT VOWELS
IN CERTAIN NON-INITIAL SYLLABLES

| | | |
|---|---|---|
| **kakkara* 'chamomile' | > | **kakkar* |
| **kakkarasta* 'chamomile.ELATIVE' | > | **kakrast* |
| Finnish *kakkara* | | Estonian *k´akar* |
| *kakkarasta* | | *k`akrast* |

(4) OBSTRUENT GEMINATION
BETWEEN A SHORT VOWEL OF AN INITIAL SYLLABLE
AND A LONG VOWEL OF A NON-INITIAL SYLLABLE

| | | |
|---|---|---|
| **käte(h)en* 'hand.ILLATIVE' | > | **kätteen* |
| Finnish *käteen* | | Estonian *k`ätte* |

incongruity. On the one hand, the changes are rather diverse as regards their typological characteristics. On the other hand, they still occurred during a relatively short time-span in an interconnected way. We treat the incongruity by claiming that the general pattern of these phonological changes is due to two subsequent purely phonetic changes, we call *the Great Rhythm Shift of Old Estonian.*

The predecessor of the modern Standard Finnish has been displaying a *syllable-switching* control of foot from time immemorial up to nowadays. However, the predecessor of the modern Standard Estonian was subject to a twofold reorganization of stress processing method before and at the time of the shift to the historic era in the Northern Europe. At the first stage of the Great Rhythm Shift, the original Balto-Finnic *syllable-switching* control of foot was abandoned in favour of an innovative *demisyllable-switching* control. At its second stage, the demisyllable-switching control was exchanged for a combination of *foot-striving and syllable-switching* control.

3.2. *From syllable-switching to demisyllable-switching.* At a time-point in the prehistory of most modern Balto-Finnic dialects, the loss of certain intervocalic obstruents, e.g. **mustata* 'black.PARTITIVE' > **mustaa*, introduced long vowels into non-word-initial syllables.

Before the change, the sound architecture of the dialects was constrained by two general principles. First, the opposition between short and long segmental vowels was possible only in a foot-initial syllable, e.g. there was a real opposition **tuli* 'fire' vs. **tuuli* 'wind' but any opposition of the type **tuli* vs. **tulii* was excluded. Second, the prototypical foot of the language was polysyllabic, e.g. the real feet **tuli* 'fire' and **tuuli* 'wind' of two syllables could not be accompanied by monosyllabic feet like **tul* or **tuul*.[5] The absolute prevalence of polysyllabic feet together with the presence of the opposition 'short vs. long segmental vowels' indicates that the Balto-Finnic dialects showed originally a syllable-switching control of foot, cf. Table 5:

Table 5
The syllable-switching control of foot
in original Balto-Finnic

| | |
|---|---|
| **ka na* | 'hen' |
| **kau na* | 'chaff/pod' |
| **kau na ta* | 'chaff/pod.PARTITIVE' |

After the change, to enable the pronunciation of the large number of words like **mustaa,* one of the original general principles had to be violated in order to retain at least the other. In Old Finnish, the prototypical foot remained polysyllabic but the opposition between short and long segmental vowels spread from foot-initial syllables to the rest of the syllables as well: **mustaa* was interpreted as a disyllabic foot with a long segmental vowel in its non-initial syllable. In Old Estonian, the opposition between short and long segmental vowels continued to be restricted to foot-initial syllables but many polysyllabic feet were replaced by monosyllabic feet so that the prototype of a polysyllabic foot vanished altogether: **mustaa* was interpreted as a string of two monosyllabic feet the latter of which contains a long vowel in its single 'foot-initial' syllable.

The option of Old Finnish, i.e. the elimination of a restriction on long vowel distribution, did not affect the conditions that determine the type of foot control and the original syllable-switching persisted, cf. Table 6:

Table 6
The syllable-switching control of foot
in Old Finnish

| | |
|---|---|
| **ka na* | 'hen' |
| **kau na* | 'chaff/pod/' |
| **kau naa* | 'chaff/pod.PARTITIVE' |

The option of Old Estonian, i.e. the introduction of the principle that monosyllabic and polysyllabic feet are prototypical to an equal extent, created conditions in which syllable-switching is impossible. In the context where the opposition between short and long segmental vowels has to be preserved (it was still necessary to differentiate between e.g. **tuli* and **tuuli*), the switching control of foot could not be exchanged for a striving control, which eliminates this opposition for sure. Consequently, the original syllable-switching control was exchanged for an innovative demisyllable-switching control, cf. Table 7 (— stands for a boundary between subsequent feet):

Table 7
The demisyllable-switching control
of foot in Old Estonian

| | |
|---|---|
| **ka na* | 'hen' |
| **kau na* | 'chaff/pod' |
| **kau - naa* | 'chaff/pod.PARTITIVE' |

The different treatment of the long vowels in words like **kaunaa* guided Old Finnish (Table 6) and Old Estonian (Table 7) to the different paths of further development as discussed above. The option of Old Finnish created a stable situation that has been persisting without greater phonetic or phonological changes for centuries. For instance, the Table 6 could as well illustrate the situation in modern Standard Finnish. However, the option of Old Estonian created an unstable situation with which a row of pervasive phonetic and phonological changes originated.

At the *beginning* of the demisyllable-switching period of Old Estonian, some phonological changes provided phonological correspondences to the phonetic adjustments that accompanied the exchange of the foot control type. In Table 4, two of such phonological changes are included.

First, the final short vowel of a trisyllabic foot was dropped, cf. (1):

(1) **kak ka ra* 'chamomile' > **kak kar*

A foot-internal syllable boundary is a very salient heuristic that points to a syllable-switching control of foot. Consequently, it is in conflict with a demisyllable-switching. The phonological change (1) is a reflection of a general phonetic process with which all trisyllabic feet (two foot-internal syllable boundaries) were eliminated so that any foot could be either monosyllabic (no boundary) or, in the extreme, disyllabic (one boundary). Note that, in this connection, the final short vowel of a disyllabic foot was retained, cf. (2):

(2) **kak ka - ras ta* 'chamomile.ELATIVE' ≡

Second, as a result of the change $h > \emptyset$, the word-internal obstruent at the beginning of a foot was geminated after a monosyllabic foot of one short syllable, cf. (3):

(3) **käte - hen* 'hand.ILLATIVE' >
**kät - teen* (rather than **kä∅ - teen*)

An obstruent is subject to an ambisyllabicity at a word-internal foot boundary between two sonorous sounds. The phonological change (3), "emergency gemination", is a reflection of the situation in which the obstruent ambisyllabicity was phonetically reinforced in order to provide monosyllabic feet of one short syllable with additional sound material. Note that the obstruent ambisyllabicity was not reinforced phonetically and did not yield gemination phonologically after monosyllabic feet ending in a long vowel, diphthong, or sonorant. In this case, the necessary sound material of a monosyllabic foot was provided by the original long syllable itself, cf. (4):

(4) **jalka - han* 'foot. ILLATIVE' >
**jal - kaan* (rather than **jalk - kaan*)

3.3. *From demisyllable-switching to foot-striving*

In a prolonged time-span perspective on a language, feet that show the same degree of accentuation tend towards an average value of their stress amount. Within the demisyllable-switching period of Old Estonian, this goal was reached by manipulating the first and the second demisyllables in certain feet. At that time, the average foot consisted of three demisyllables. To approximate to its average stress amount, feet consisting of two or four demisyllables were altered. In two-demisyllable feet, the second demisyllable was reinforced, we designate the reinforcement by ¯ (it was inconvenient to reinforce the first, foot-initial demisyllable that had a reinforced value already by itself); in four-demisyllable feet, the first demisyllable was reduced, we designate the reduction by ˘ (Table 8).

Table 8
The trend towards the average stress value of feet in Old Estonian

| | | | | |
|---|---|---|---|---|
| **ka na* | 'hen' | 2 ds | + | *[kanā] |
| **ka nan* | 'id.GENITIVE' | 3 ds | OK! | *[kanan] |
| **kau na* | 'chaff/pod' | 3 ds | OK! | *[kauna] |
| **kau nan* | 'id.GENITIVE' | 4 ds | – | *[kăunan] |
| **kau – (naa)* | 'id.PARTITIVE' | 2 ds | + | *[kaū] |
| **kau – (naan)* | 'id.ILLATIVE' | 2 ds | + | |
| **maan* | 'earth.GENITIVE' | 3 ds | OK! | *[maan] |

The trend towards the average stress value of all feet was a statistical tendency that manifested itself clearly in long speech stretches. However, language users reduced it to a single phonetic rule that was applicable to every concrete foot. Because of the trend, both in two-demisyllable and four-demisyllable feet, the energetical value of the first two demisyllables became roughly equal, they revealed more or less the same amount of stress. Relying on this correspondence, language users deduced a general phonetic principle, *demisyllable balancing*, according to which the first two demisyllables of all feet have to display the same amount of stress. As for two-demisyllable and four-demisyllable feet, the rule was applicable without complications. However, as for three-demisyllable feet, it caused additional phonetic adjustments: in order to raise the energetical value of the second demisyllable to the level of the first demisyllable, some of the stress of the third demisyllable had to be reattached to the second demisyllable instead (Table 9).

The phonetic adjustments of the three-demisyllable feet in the context of demisyllable balancing triggered a chain of phonetic and corresponding phonological changes that exerted a crucial influence on speech rhythm. As the final result, they created conditions in which the demisyllable-switching control of foot was exchanged for a combination of foot-striving and syllable-switching control. In other words, they introduced the *ending* of the demisyllable-switching period of Old Estonian. In Table 4, this chain is represented by the phonological performance of its three main changes.

First, the final short vowel of a disyllabic three-demisyllable foot was dropped, cf. (5):

(5) **kak ka – ras ta* 'chamomile.ELATIVE' > **kakk – rast*

Table 9
The application of demisyllable balancing (the first two demisyllables of a foot display the same amount of stress) in Old Estonian (⇑ stands for a manipulated command)

2-demisyllable feet

**ka na* > *[kanā], **kau* > *[kaū]

1st↳ ≈ 2nd⇑↳ : OK!

4-demisyllable feet

**kau nan* > *[kăunan]

1st⇑↳ ≈ 2nd↳ : OK!

3-demisyllable feet

**ka nan* > *[kanan], **kau na* > *[kauna],
**maan* > *[maan]

1st↳ ≠ 2nd↳ :
<u>additional phonetic adjustments</u>

*[kanan] > *[kanān̆], *[kauna] > *[kāună]
*[maan] > *[maān̆]

The result: 1st↳ ≈ 2nd⇑↳

This phonological change reflects the phonetic leveling of a consonant-ending foot like *[maān̆] and a vowel-ending foot like *[kāună] by the shortening process *[kāună] > *[kāūn^a] > *[kāūn̆].

The above phonetic leveling ushered in an *overall phonetic shortening* of vowels in non-word-initial syllables. The phonetic process *[kāună] > *[kāūn^a] was extended, for instance, to the final short vowel of a disyllabic two-demisyllable foot, e.g. to the second [a] in *[kanā]. However, here a real phonological loss appeared only in tonic non-word-initial feet; in accentuated word-initial feet the vowel persisted, cf. (6):

(6) **ka va – la ta* 'sly.PARTITIVE' > **kava – lat*

Second, long vowels in all non-word-initial syllables were shortened, cf. (7):

(7) **mus – taa* 'black.PARTITIVE' > **mus ta*

This phonological change reflects the same overall phonetic shortening of vowels in non-word-initial syllables as in the previous case: on the analogy of e.g. *[kāună] > *[kāūn^a] > *[kāūn̆], the phonetic process *[kaū – naā] > *[kaū – naa] > *[kāuna] started. Note that the shortening of long vowels replaced the original word-internal sequences of monosyllabic feet by innovative single disyllabic feet. In the original *[kaū-naā], the second syllable consisted of two demisyllables and, thus, could form a separate foot. In the innovative *[kāuna], the second syllable consisted of one demisyllable only and, thus, had to cohere with the preceding foot.

Third, and that is the crucial point, the opposition of two contrastive accents, the even accent (´) and the sharp accent (`), arose, cf. (8):

(8) **kau nan* 'chaff/pod.GENITIVE' **k´au nan*
**kau – naan* 'chaff/pod. ILLATIVE' **k`au nan*

This phonological change was an immediate functional result of the fact that the demisyllable structure of the innovative disyllabic feet (**kaunan* from **kau-naan*) merged often with that of some original disyllabic feet (**kaunan* from **kaunan*). This merger was functionally inconvenient, as it could give rise to a large-scale homonymy, cf. the difference of the grammatical meaning in **kaunan* 'chaff/pod.GENITIVE' vs. **kau-naan* 'chaff/pod.ILLATIVE' and of the lexical meaning in **kiiren* 'ray.GENITIVE' vs. **kii-reen* 'quick.GENITIVE'. In this context, the original phonological contrast between short and long segmental vowels in non-word-initial syllables (**a* vs. **aa*) was reinterpreted as a phonological contrast between two different types of feet (**k´aunan* vs. **k`aunan*) relying on those of their phonetic properties that had appeared differently in connection with the short and long vowels before the change.

The two types of disyllabic feet differed in the behaviour of the energy level of their first, second, and third demisyllable taken as an integral pattern. In feet like **kaunan* from **kaunan*, the demisyllables displayed an *even* energy level of a low value throughout the pattern, as it had been characteristic of the original 4-demisyllable feet. In feet like **kaunan* from **kau-naan*, the first two demisyllables displayed an even energy level of a high value, as it had been characteristic of the original 2-demisyllable feet but the third, shortened demisyllable displayed an energy level of much lower value. Correspondingly, there was a pattern of a broken energy level with a *sharp* projection at the boundary of the second and third demisyllables. The two different energetical patterns that resulted from different combinations of demisyllables were reanalyzed as two different energetical patterns that could characterize feet with one and the same demisyllable composition. After the reanalysis, the even pattern stood for the *even accent* characterized by an even energy distribution in the foot-initial syllable and a smooth beginning of the subsequent foot-internal syllable. Note that in the process of total accent split, the even accent was found not only in feet like **k´aunan* 'chaff/pod.GENITIVE' but also in feet like **k´ana* 'hen' displaying a foot-initial short syllable. On the contrary, the sharp pattern stood for the *sharp accent*, characterized by an localized energy distribution in the foot initial syllable, i.e. a sharp movement to the maximal energy level at its end, and an apart beginning of the subsequent syllable. Note that in the process of total accent split, the sharp accent was found not only in feet like **k`aunan* 'chaff/pod.ILLATIVE' but also in feet like **k`aun* 'chaff/pod' of one syllable (Table 10).

Table 10
The accent split of Old Estonian

| | | | |
|---|---|---|---|
| 1. even pattern | | > | even accent |
| **kau nan* | 'chaff/pod.GENITIVE' | > | **k´au nan* |
| **ka na* | 'hen' | > | **k´a na* |
| 2. broken pattern with a sharp projection | | > | sharp accent |
| **kau nan* | 'chaff/pod.ILLATIVE' | > | **k`au nan* |
| **kaun* | 'chaff/pod' | > | **k`aun* |

It was a rise of contrastive accents that moved Old Estonian from the type of demisyllable-switching languages to the complex type in which foot-striving and syllable-switching are interwoven. As for striving, contrastive accents entail the manipulation of the maximal energy region in a foot that is out of the question in pure switching languages. As for foot-striving, there were polysyllabic feet in Old Estonian that excluded the possibility of syllable-striving. As for switching, the opposition between short and long segmental vowels had still to be preserved (it was necessary to differentiate between e.g. **t´ulen* 'fire.GENITIVE' and **t´uulen* 'wind.GENITIVE'), so striving commands that eliminate the distinction for sure did not exchange the original switching commands entirely. As for syllable-switching, the equal prototypicality of polysyllabic and monosyllabic feet was no more a hindrance: monosyllabic feet like **k`aun* could be handled by a combination of a switching and a striving command rather than needed two subsequent switching commands in one syllable.

We have shown that the pervasive changes in the phonological structure of Old Estonian during the first centuries of the second millenium A.D. were accompanied by an equally thorough-going revision in the speech rhythm appearance of the language. Note that the path from syllable-switching via demisyllable-switching to foot-striving had to be especially manifest in the history of the Insular dialect of modern Estonian. Here, the sharp accent appears in the form of a *pure* striving command, cf. (9):

(9) *k`ɔuna* 'pod.GENITIVE'

*k`ɔunə* 'pod.PARTITIVE'

## 4. SOUND CHANGE PRESCRIPTION

According to Lass 1980, no inherent explanation is available for any linguistic change: he differentiates, inter alia, between deductive-nomological explanations and teleological explanations and argues against the possibility of either in the context of the diachronic research. On the contrary, Itkonen 1986 argues for the possibility of *teleological explanations* of linguistic changes: according to him, any linguistic change may be seen as contributing somehow to the increase of the form-meaning isomorphism in linguistic units. Still, for him, deductive-nomological explanations of linguistic changes are out of the question. We try to show that, in the framework where phonological-phonetic and purely phonetic sound changes are kept apart, the possibility of *deductive-nomological explanations* of linguistic changes cannot be excluded categorically.

A sound change may allow two types of formalizations with regard to formalization strength. A *description* represents it as a transformation that *occurred* given certain conditions, a *prescription* represents it as a transformation that *has to occur* given certain conditions. Prescription is preferable to description. Standing on a higher level of abstraction, it fits in with all functions a formalization of a sound change may serve rather than is appropriate for single explicit aims only. As a matter of fact, a deductive-nomological, "law-like" explanation of a sound change is equal to its prescription.

In order to provide a sound change with a prescriptive formalization, it is inevitable to extract exhaustively and arrange in a mutually exclusive, categorical manner all the *factors* that trigger the sound change in question. We have presented phonology as a social phenomenon and phonetics as a psycho-physiological and acoustic phenomenon. In this context, the prescription of a *phonological* change is precluded in principle: the sociolinguistic constraints on sound changes are many and fuzzy to the extent that any exhaustive factor extraction/arrangement is an insoluble task for a human being. However, some *phonetic* changes could allow a prescriptive formalization: the physiological and environmental constraints on sound changes are few and distinct to the extent that the needed exhaustiveness in factor extraction and arrangement may be achievable at least in some cases.

We have proposed that two different kind of sound changes should be extracted in the course of the sound history of a concrete language. As for *phonological-phonetic* changes, the above reasoning excludes prescription in principle; as for *purely phonetic* changes, it leaves the possibility open.

The essence of this claim is exemplified by comparing the actual course of the Great Rhythm Shift, i.e. a row of purely phonetic changes, and corresponding phonological-phonetic changes in the history of two different Balto-Finnic dialects, i.e. Standard Estonian and Southwestern Finnish.

## 5. THE GREAT RHYTHM SHIFT: THE SOUTH-WEST OF THE BALTO-FINNIC AREA

Up to here, we have treated the Great Rhythm Shift as present in the history of Estonian but absent in the history of Finnish. However, the straightfoward distinction crumbles if to consider the sound history of Estonian and Finnish dialects in addition to the history of standard languages. On the one hand, the conservative development has to be extended from Finnish alone to Northeastern Estonian also; on the other hand, the Great Rhythm Shift has to be extended from Estonian alone to Southwestern Finnish also (Map 1).


Map 1. The Great Rhythm Shift in Balto-Finnic (to the southwest of the line ).

The presence of the Great Rhythm Shift both in the history of Standard Estonian and Southwestern Finnish provides a rather unique case of a row of phonetic and phonological changes that applied to the essentially same sound material (the original similarity of Balto-Finnic dialects) in different communities (the opposite coasts of the Gulf of Finland) not affected by pervasive mutual contacts (cf. the opinion expressed in the articles of the representative collection Gallén 1984). As a matter of fact, around 1000 A.D. both Estonian and Southwestern Finnish were influenced rather by Old Norse, the language of the Vikings. We follow Wiik 1986 in claiming that the Scandinavian vowel balancing could affect the speakers of the Balto-Finnic dialects in question: when the critical words like **mustaa* appeared, they followed the demisyllabic-switching option that was already known from the speech of the foreigners.



In this context, the exact comparison of the Estonian and Southwestern Finnish sound changes may display far-reaching theoretical implications.

On the one hand, the possibility of a prescriptive formalization of some purely phonetic changes is supported by the fact that the purely phonetic course of the Great Rhythm Shift had to be rather identical in both dialects. On the other hand, the fundamental impossibility of a prescriptive formalization of phonological-phonetic changes is supported by the fact that the phonological extensions of the Great Rhythm Shift did not coincide in the Estonian and Southwestern Finnish norm.

Indeed, some phonologizations could be the same for the two dialects. For instance, the Estonian and Southwestern Finnish patterns of vowel shortenings in non-initial syllables are rather similar. However, and that is the main point, the same phonetic changes could lead to rather different phonologizations as well. For instance, the counterpart of the Estonian "emergency" gemination in **kätehen* 'hand.ILLATIVE' > **kät-teen* (cf. the lack of gemination in **jalkahan* 'foot.ILLATIVE' > **jal-kaan*) was the Southwestern Finnish pervasive "special" gemination both in **kätehen* > **kät-teen* and **jalkahan* > **jalk-kaan*. Similarly, the shortening of long vowels in non-word-initial syllables triggered in Estonian the opposition of contrastive accents (**kaunan* 'chaff/pod.GENITIVE' > **k´aunan* vs. **kau-naan* 'chaff/pod.ILLATIVE' > **k`aunan*) but yielded a total merger in Southwestern Finnish (**kaunan*, **kau-naan* > **kaunan*). Note that, because of the latter difference, Estonian has turned into a foot-striving language but Southwestern Finnish still continues as a demisyllable-switching language.

## NOTES

1. In addition, speech flow has the function to convey information about the person who is speaking (cf. Ladefoged 1984: 84) and about the extralinguistic, pragmatic situation in which he is speaking. Note, however, that these are not the special functions of speech flow, they are rather the functions of whatever human sound, be it e.g. speech, cry or wheeze.

2. We are aware of the bad connotation of the term 'speech energy'. In general, energy is something to be measured. Indeed, the exact amount of acoustic energy may be computed easily from the temporal intensity envelope for different frequency bands. However, the fixation of the exact value of the physiological energy while speaking or listening is a too complex task to be solved by exact measurements nowadays. Here, the proper term to indicate the generalization degree we mean would be rather 'the presence of muscular vs. neural activity' rather than 'physiological energy'. Nevertheless, the cover term 'physiological energy' has to be used in order to point to the fact that speaker's physiological activities, acoustic energy, and listener's physiological activities form a unitary chain. As a matter of fact, acoustic energy cannot be dealt with as a physiological activity.

3. In psycholinguistic literature, the term pair 'accentuation vs. deaccentuation' stands for our 'maximal vs. minimal accentuation'. We emphasize the gradual rather than the directional nature of the phenomenon.

4. Our account on stress-timing differs from that of Dauer 1983 as being based on a purely phonetic argumentation. Relying on the postulate that phonetics and phonology are incompatible within one treatment, we deduce the scale nature of stress-timing from the diversity of the possible patterns in which stress beats may be revealed rather than from an interplay of phonological, phonetic, lexical, and syntactic facts about a particular language.

5. In original Balto-Finnic, there was a handful of monosyllabic feet in addition, e.g. **maa* 'earth'. However, they could not affect the general prototype of foot polysyllabicity.

## REFERENCES

D. Abercrombie 1967. Elements of General Phonetics. Edinburgh: Edinburgh University Press.
G.D. Allen 1975. Speech rhythm: its relation to performance universals and articulatory timing. - Journal of Phonetics 3: 75 - 86.
H. Andersen 1973. Abductive and deductive change. - Language 49, 4: 765 - 793.
R. Anttila 1972. An Introduction to Historical and Comparative Linguistics. New York: Macmillan.
P. Ariste 1938. A quantitative language. - Proceedings of the Third International Congress of Phonetic Sciences: 276 - 280. Ghent.
T. Balasubramanian 1979. Timing in Tamil. - Edinburgh University, Department of Linguistics, Work in Progress 12: 57 - 74.
P.M. Bertinetto 1981. Strutture prosodiche dell'italiano. Firenze: Accademia della Crusca.
L. Biedrzycki 1980. Isochronous feet in a reading of Polish verse. - Edinburgh University, Department of Linguistics, Work in Progress 13: 50 - 59.

B. Comrie 1981. Language Universals and Linguistic Typology. Oxford: Basic Blackwell.

R.M. Dauer 1983. Stress-timing and syllable-timing reanalyzed. - Journal of Phonetics 11, 1: 51 - 62.

A. Eek 1974. Observations on the duration of some word structures: I. - Estonian Papers in Phonetics 1974: 16 - 30.

A. Eek 1982. Stress and associated phenomena: a survey with examples from Estonian. I. - Estonian Papers in Phonetics 1980-1981: 20 - 59.

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

A. Eek, M. Remmel 1974. Context, contacts and duration: two results concerning temporal organization. - Speech Communication Seminar, Speech Production and Synthesis by Rules, 2: 187 - 192. Stockholm.

C.A. Fowler 1979. 'Perceptual centers' in speech production and perception. - Perception and Psychophysics 25: 375 - 388.

C.A. Fowler 1981. A relationship between coarticulation and compensatory shortening. - Phonetica 38: 35 - 40.

O. Fujimura 1983. A linear model of speech timing. - Paper to be published in: In Honor of Ilse Lehiste, R. Channon, L. Shockey, eds.

O. Fujimura, J.B. Lovins 1978. Syllables as concatenative phonetic units. - Syllables and Segments: 107 - 120. A. Bell, J.B. Hooper, eds. North-Holland Publishing Company.

H. Fujisaki, N. Higuchi 1979. Temporal organization of segmental features in Japanese disyllables. - University of Tokyo, Research Institute of Logopedics and Phoniatrics 13: 155 - 161.

J. Gallén (ed.) 1984. Suomen väestön esihistorialliset juuret. - Bidrag till kännedom av Finlands natur och folk. Finska Vetenskaps-Societeten, 131. Helsinki.

R.E. Hammarberg 1976. The metaphysis of coarticulation. - Journal of Phonetics 4: 353-363.

Y. Homma 1981. Durational relationship between Japanese stops and vowels. — Journal of Phonetics 9: 273 - 281.

P. Howell 1984. An acoustic determinant of perceived and produced anisochrony. - Proceedings of the Tenth International Congress of Phonetic Sciences, vol. IIB: 429 - 433.

E. Itkonen 1986. Form-meaning isomorphism, or iconicity, in diachronic linguistics (and elsewhere). - Symposium on Formalization in Historical Linguistics, Tallinn, November 1986, Summaries: 38 - 46.

D.H. Klatt 1973. Interaction between two factors that influence vowel duration. - Journal of the Acoustical Society of America 54: 1102 - 1104.

P. Ladefoged 1968. Linguistic aspects of respiratory phenomena. - Annals of the New York Academy of Sciences 155: 141 - 151.

P. Ladefoged 1980. What are linguistic sounds made of? - Language 56, 3: 485 - 502.

P. Ladefoged 1984. 'Out of Chaos Comes Order'; physical, biological, and structural patterns in phonetics. - Proceedings of the Tenth International Congress of Phonetic Sciences, IIB: 83 - 95.

P. Ladefoged, J.D. De Clerk, M. Lindau, G. Papcun 1972. An auditory-motor theory of speech production. - UCLA Working Papers in Phonetics 22: 48 - 75.

R. Lass 1980. On Explaining Language Change. Cambridge: University Press.

W.A. Lea 1974. Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding. - Univac Report No. PX10791. St. Paul, Minnesota: Sperry Univac, DSD.

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

I. Lehiste 1977. Isochrony reconsidered. - Journal of Phonetics 5: 253 - 263.

J. Lehtonen 1970. Aspects of quantity in Standard Finnish. - Studia Jyväskyläensia 6. Jyväskylä.

A.M. Liberman, F.S. Cooper, K.S. Harris, P.F. MacNeilage 1962. A motor theory of speech perception. - Proceedings of the Speech Communication Seminar, vol. 2: Paper D 3, Stockholm.

B. Lindblom 1963. Spectrographic study of vowel reduction. - Journal of the Acoustical Society of America 35: 1773 - 1781.

B. Lindblom, K. Rapp 1973. Some temporal regularities of spoken Swedish. - PILUS 21: 1-59.

P.F. MacNeilage 1972. Speech physiology. - Paper prepared for the conference on Speech Production and Perception; Their Relation to Cortical Function, Vancouver, April 12-14, 1972.

S.M. Marcus 1981. Acoustic determinants of perceptual center (P-center) location. - Perception and Psychophysics 30, 3: 247 - 256.

L. Menn, B. MacWhinney 1984. The repeated morph constraint: toward an explanation. - Language 60, 3: 519 - 541.







P. Mermelstein 1975. Automatic segmentation of speech into syllabic units. — Journal of the Acoustical Society of America 58: 880 - 883.

M. Miller 1984. On the perception of rhythm. - Journal of Phonetics 12: 75 - 83.

T. Navarro 1932. Manual de Pronunciación Española. Centro de Estudios Históricos, Madrid.

S.G. Nooteboom 1972. Production and Perception of Vowel Duration. A Study of Durational Properties of Vowels in Dutch. Utrecht.

J.D. O'Connor 1973. Phonetics. Harmondsworth, Middlesex: Penguin Books.

J.J. Ohala, C.J. Riordan, H. Kawasaki 1979. Investigation of pulmonic activity in speech. - Proceedings of the Ninth International Congress of Phonetic Sciences, vol. 1: 205.

K.L. Pike 1946. The Intonation of American English. Ann Arbor: University of Michigan Press.

R. Plomp 1984. Perception of speech as a modulated signal. - Proceedings of the Tenth International Congress of Phonetic Sciences, II B: 29 - 40.

P. Roach 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. - Linguistic Controversies: 73 - 79, D. Crystal, ed., London.

M. Sawashima, H. Hirose, H. Yoshioka 1982. Interaction between articulatory movements and vocal pitch control in Japanese word accent. - University of Tokyo, Research Institute of Logopedics and Phoniatrics, Annual Bulletin 16: 11 - 19.

D.R. Scott 1982. Duration as a cue to the perception of a phrase boundary. — Journal of the Acoustical Society of America 71: 996 - 1007.

D.R. Scott, S.D. Isard, B. Boysson-Bardies 1985. Perceptual isochrony in English and in French. - Journal of Phonetics 13: 155 - 162.

J.M.B. Terken 1984. The effect of accentuation on comprehension: an experiment. — Proceedings of the Tenth International Congress of Phonetic Sciences, II B: 558 - 561.

M. Vayra, C. Avesani, C.A. Fowler 1984. Patterns of temporal compression in spoken Italian. - Proceedings of the Tenth International Congress of Phonetic Scienes, II B: 541-546.

K. Wiik 1965. Finnish and English Vowels. A Comparison with Special Reference to the Learning Problems Met by Native Speakers of Finnish Learning English. Annales Universitatis Turkuensis B, 94. Turku.

K. Wiik 1986. Suomi ja sen lähimmät sukukielet kahden germaanisen äännelain puristuksessa. - Paper presented at Institute of Language and Literature, Tallinn, April 4, 1986.

Л.В. Бондарко, Л.А. Вербицкая, Л.Р. Зиндер 1966. Акустическая характеристика безударности (на материале русского языка). - Структурная типология языков. Москва.

Л.Р. Зиндер 1964. Влияние темпа речи на образование отдельных звуков. — Ученые записки Ленинградского университета 325, Серия филологических наук 69: 3 - 27.

В.Б. Касевич 1983. Фонологические проблемы общего и восточного языкознания. Москва: Наука.

Л.А. Чистович, В.А. Кожевников et al. 1965. Речь. Артикуляция и восприятие. Москва, Ленинград.