Estonian and English rhythm:
a two-dimensional quantification based on syllables and feet

Eva Liina Asu[1] & Francis Nolan[2]

[1]Institute of the Estonian Language, Tallinn, Estonia
[2]Department of Linguistics, University of Cambridge, UK
`ela@cantab.net; fjn1@cam.ac.uk`

Abstract

This paper expands a recent pilot experiment [3] on Estonian rhythm within the quantificational approach to the study of rhythm, using the Pairwise Variability Index (PVI). The PVI expresses the average difference between adjacent phonological units such as vowels, consonantal intervals or syllables. It is argued here that confining the application of the PVI to the level of the syllable (or its components) misses the essence of Estonian rhythm and indeed of phonetic rhythm in general, and the first experiment reported in this paper quantifies Estonian rhythm in terms of the durational PVI of both the syllable and (innovatively) the foot. In the second experiment, results are compared with the same measures for another language with strong stress, English. Both languages have a similar, relatively low foot PVI, but English has a considerably higher syllable PVI reflecting its radical reduction of unstressed syllables in polysyllabic feet.

1. Introduction

The Pairwise Variability Index (PVI) is a metric used for quantifying speech rhythm. It measures the average variability of a property, usually duration, from one unit to the next. It has most commonly been used to express the durational patterning of successive vowels or successive intervocalic (consonantal) intervals, showing how each linguistic event differs from the next. The PVI provides an alternative to the traditional view of absolute isochrony ('syllable timing' vs. 'stress timing' [1]), implying instead a scalar 'prominence gradient' between successive units. The research reported in e.g. [11] and [8], focusing on the PVI, and in e.g. [14], using different quantitative measures, has shown that it is possible to achieve useful scalar characterisation (but not discrete categorisation) of the rhythm of different languages.

The PVI was first applied in a study of Singapore English rhythm [10] where it was demonstrated that Singapore English had a lower average vocalic PVI than British English, confirming earlier impressionistic observations about Singapore English being more 'syllable timed' than British English, a prototypically 'stress timed' variety.

Estonian, according to the traditional rhythmic dichotomy, is classified as a syllable timed language [7] but is also said to be characterised by foot isochrony [15]. Main stress in Estonian words is fixed on the first syllable which, together with the following unstressed syllable, constitutes the domain of the three-way quantity contrast for which Estonian is famous. Estonian was included in a rhythmic comparison of a number of languages [8] that were quantified using the PVI for the duration of successive vowels and, as an independent dimension, the PVI of successive intervocalic intervals. Estonian showed a vocalic variability roughly in the same range as French, Catalan, Rumanian, and Polish. According to a separate study [16], the vocalic PVI for Latvian (traditionally also a syllable timed language) was shown to be very similar to that of Estonian in [8].

There are, however, two curious aspects to the PVI research tradition. The first is why, when 'syllable-timing' is at issue (i.e. the tendency of a language to make syllables the same length), the pairwise variability not of the syllable but of components of the syllable (vowel and intervocalic interval) has been favoured (except in [4]). Low [10] attributes her choice of the vocalic PVI to Taylor [17], who claims that vowel duration is the key to syllable timing. In practical terms this choice also allows researchers to side-step controversies about English syllable-division, but little detailed justification can be found in the literature. Subsequently, the intervocalic (CPVI) measure was adopted to capture, in particular, languages' variability in permitted consonant sequences. In principle, however, neither of these PVI measures seems entirely appropriate to capture the notion of 'syllable-timing'.

The second curious aspect, given the conventional opposing term 'stress-timing' (i.e. tendency of a language to compress syllables to yield isochronous feet), is that little if any attention has previously been paid to the PVI of the foot. Ramus [13] suggests the foot as a possible alternative unit for the measure of speech rate but as far as the present authors are aware the foot PVI has not been calculated in any language. A reason for this neglect may have been an assumption that quantifying syllable features exhausts rhythmic characterisation, i.e. that if a language has a low vocalic PVI it is syllable timed, and if it has a high PVI it is stress timed. It is this assumption that will be challenged here.

The research carried out for the present paper builds on a pilot experiment reported in [3]. It extends the Estonian dataset used for that study, and provides results for an equivalent sample of English speakers as a comparison. PVI measures are calculated for five different units. On the basis of the results it will be argued that syllable-based PVI measures are logically independent of the foot PVI, and that the rhythmic nature of languages (at least those with identifiable stress-feet) is more completely captured by looking at both syllables and feet.

2. Experiment I: Estonian PVI measures

2.1. Materials and subjects

The materials consisted of a read passage originally recorded for the analysis of intonation [2]. The text used for the present purposes comprised 178 syllables.

The data from five young female speakers of Standard Estonian was analysed. Three subjects were recorded in a

quiet environment in Tartu, Estonia, and two in the sound-treated booth of the phonetics laboratory of Cambridge University. The subjects were asked to read the passage at a normal tempo. Depending on the speaker, the recorded material yields 12 to 22 intonation phrases.

2.2. Analysis

For the Estonian data, the same two sets of measurements were made as in [3]. First, the start times of each vowel and of each intervocalic interval were measured. The segmentation of glides such as [j] is highly problematic but was nonetheless attempted for completeness on the basis of formant dynamics and careful listening. The start of the vowel preceded by a stop consonant was measured from the burst of the consonant rather than the beginning of the first formant. From the vocalic and intervocalic measurements the vowel PVI (VPVI) and intervocalic PVI (CVPI) were calculated.

Additionally, the vocalic and intervocalic measurements allowed the derivation of 'pseudo-syllables' i.e. units consisting of an intervocalic interval and the following vowel [4], [12]. The motivations for calculating the pseudo-syllable include the fact that it is the natural corollary of the 'vocalic-intervocalic' PVI dichotomy, and, more interestingly, that it corresponds to the so called 'Articulatory Syllable' [9], which was proposed as the domain of coarticulation. Although many studies disconfirmed this hypothesis, the pseudo-syllable nevertheless has a heritage in research into the organisation of speech production which makes it worth considering. In addition Pellegrino et al. [12] show, for instance, that the pseudo-syllable is effective in automatic language discrimination.

The second set of measurements took as its starting point a traditional phonological syllabification of the utterances. There is relatively little controversy over how Estonian (unlike English) syllabifies: e.g. a long (Q2) or overlong (Q3) vowel or a diphthong forms one syllable but consonant clusters of two or more consonants are split so that the last consonant starts a new syllable. Acoustically some arbitrary decisions had to be made; for instance long consonants, or sequences of identical vowels at the end of one word and the beginning of another, were simply divided at their mid-point; while sequences of two different vowels at word boundaries were divided at the point which best preserved their acoustic and auditory identity. Most cases, however, were unproblematic. The beginning of each syllable was recorded, and the syllable lengths calculated.

A further set of durations was derived from the linguistic syllable, namely those of phonological feet. These are considered to consist of a stressed (not necessarily accented) syllable and zero, one, or two following unstressed syllables. Trisyllabic words constitute one prosodic foot if there is no secondary stress on the second or third syllable of the word [15]. Phrase-initial unstressed syllables (an 'anacrusis') [5] were left out as they do not participate in a well-formed foot. The calculation of foot durations involved adding together the durations of the syllables making up the foot.

The PVI can be calculated in two ways: 'raw' (rPVI), where the differences between successive pairs of units are averaged over the material, or 'normalised' (nPVI). Normalisation involves expressing each difference as a proportion of the average of the two units involved (e.g. their average duration). The original point of this [10] was to neutralise the effect of utterance level rate variation, particularly between-speaker differences in rate and phrase final lengthening. There are arguments for and against this (e.g. [4], [8]) as a matter of principle, but the fact that our units are of widely differing size (segment, syllable, foot) means that normalisation is essential. The magnitude of equivalent variation between feet will inevitably be greater in absolute terms than that between syllable-parts, but expressing the variation as a proportion of the two units involved neutralises this difference of magnitude. The resultant fractional value of each normalised PVI is multiplied by 100 to express it as a percentage.

2.3. Results

Figure 1 summarises the results for Estonian, presenting five normalised PVI measures and five speakers within each measure. The first measure is the normalised vowel PVI (nVPVI). The average value of all speakers is 48.3, which compares well to 44.6 from the earlier pilot study [3] and that of Grabe and Low's 45.4 [8].
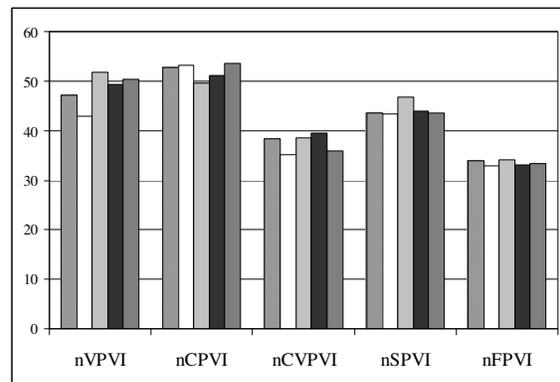

Figure 1: *Estonian normalised PVI measures for five speakers.*

The second measure plotted in Figure 1 is the normalised intervocalic PVI (nCPVI), the mean of the speakers being 52.0. This measure was not calculated in Grabe and Low's study where instead the raw intervocalic PVI was used. Our group mean of 40.8 (not shown in the figure as it is not comparably scaled) is very close to 40.0 in [8], despite not being normalised and therefore very sensitive to speech rate.

The next two PVI measures in Figure 1 are those representing the syllable: 37.5 for the mean PVI of the 'pseudo-syllable' comprising a vowel and all preceding consonantal material (nCVPVI), and 44.2 for the linguistic syllable (nSPVI). Finally, the foot PVI (nFPVI) values for the five speakers are presented having the lowest group mean at 33.5.

A paired samples t-test shows that nFPVI is different from each of nVPVI, nSPVI and nCVPVI at the $p < 0.01$ level. The vocalic measure (nVPVI) is not significantly different from the pseudo-syllable measure (nCVPVI), to which it presumably contributes substantially, but it is significantly different from the linguistic syllable (nSPVI) at the $p < 0.01$ level.

2.4. Discussion

This suggests, as a minor finding, that caution may be needed in equating the vocalic interval with 'real' syllable timing. The present data give us other reasons to question the desirability of splitting the syllable. First, nVPVI has the least between-speaker stability (SD around 3.5, compared to 1.4 for the linguistic syllable measure and 0.5 for the nFPVI). This suggests that timing regularity below the syllable may not be controlled as tightly as higher up the prosodic hierarchy. Second, the consonantal measure nCPVI reflects two very different influences: the degree of phonotactic permissiveness, and the dynamics of individual consonants (in Estonian, for instance, the presence of a very short tapped /r/ at one extreme and 'overlong' (Q3) consonants at the other).

However, the most important discovery is that the foot measure nFPVI turns out to be significantly smaller than our syllable-based measures (nCVPVI, nSPVI, and, arguably a syllable-based measure, nVPVI). This confirms a greater tendency (albeit far from complete) towards foot- than syllable isochrony [15], despite the patterning of Estonian with syllable timed languages on syllable-based measures.

Across languages, we would expect nFPVI to have a variable relationship with syllable-based measures, a relationship which will define rhythmic type in a rather subtle way. To test this, we carried out a comparison with English. Specifically, for English, we predict a larger syllable PVI (nSPVI) than for Estonian because of the trend in English (unlike Estonian) towards radical reduction of unstressed vowels.

3. Experiment II: comparison with English

3.1. Materials and subjects

For the comparison with English, part of the reading passage (the 'Cinderella' story) from the IViE corpus [18] was used. The read sample consisted of approximately 157 syllables underlyingly.

The subjects were five female speakers of Standard Southern British English from Cambridge aged between 16 and 19. They were recorded in a quiet room in a local school as part of the data collection for the IViE project.

3.2. Analysis

Since the main interest arising from the analysis of Estonian is the relationship between the syllable and foot PVIs, only these were measured for English The demarcation of syllables is more problematic in English than Estonian. This is partly because English allows more complex consonant sequences, and partly because English has been the target of more phonologists. Generally, CV syllabification was maximised within words, but syllables were not allowed to straddle word boundaries; thus *they . dreamed . of . we.dding . bells*.

Feet are also more problematic in English. There are two broad approaches, the more modern suggesting that every non-reduced vowel constitutes a foot (so that 'modest gymnasts' [mɒdəst dʒɪmnæsts] consists of three feet), and the more traditional requiring a rhythmic stress at the start of each foot (so that 'modest gymnasts' consists of two feet). The former ('full vowel') strategy was chosen here, first because the determination of rhythmic stresses is difficult, and second

because the 'full vowel' foot has been shown to be perceptually important in lexical access [6].

3.3. Results

This study of English appears to confirm our prediction. Figure 2 compares mean values of the normalised syllable and foot PVIs for the two languages with the grey bars representing Estonian and the black bars English. It can be seen that the mean nSPVI is higher in English (59.1) than in Estonian (44.2). An unpaired t-test showed this between-language difference to be significant at the p<0.01 level. On the other hand, the average foot PVI (nFPVI) was virtually identical (though with a larger SD) and the difference between the languages non-significant: 33.9 for English as compared to 33.5 for Estonian.
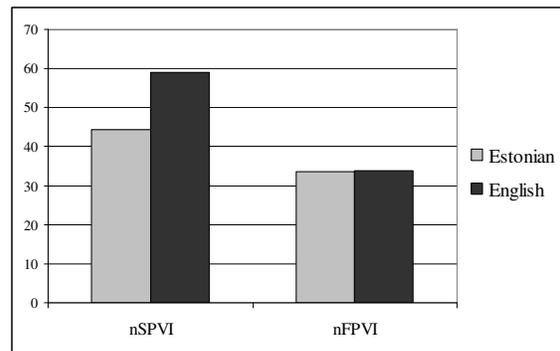

Figure 2: *Comparison of Estonian and English average normalised syllable and foot PVIs.*

3.4. Discussion

This result strongly indicates that both languages strive towards foot-isochrony, but English employs a strategy of unstressed vowel reduction which boosts its syllable PVI. We infer that in Estonian, as a foot accrues unstressed syllables, compensatory shortening is shared more equally among the syllables including the stressed syllable.

4. General discussion

The results presented above suggest that languages need not, as in the traditional dichotomy, either (like English) squash their unstressed syllables to achieve approximate foot-isochrony, or (like French) keep their syllables fairly even and not bother about foot timing. They could also equalise their feet to some degree, but share the 'squashing' more democratically in polysyllabic feet. Estonian, with its strong stress but near absence of vowel quality reduction in unstressed syllables, and despite its three-way quantity contrast which sporadically curtails syllable-equality, may be at base such a language.

To use a musical analogy, both Estonian and English might have the same time signature (e.g. 3/8), but differ in that Estonian is more inclined to have three quavers (eighth notes) in a bar but English a crotchet and two semiquavers:

Of course, language is not rhythmically consistent like music, and successive 'bars' are highly variable, but the notion that

two equally strongly rhythmical languages could differ by one (i.e. English) having a more 'lilting' rhythm is not implausible.

Given orthogonal nFPVI (foot) and nSPVI (syllable) measures, a truly syllable timed language would have high nFPVI and low nSPVI. The problem may be whether such languages have identifiable feet. The fourth logical possibility of high nFPVI and high nSPVI might not occur, since it would imply a rather perverse adjustment of syllables in a direction contrary to that which would achieve foot isochrony.

We are aware that the experiments reported raise a number of thorny problems. In particular, defining both syllables and feet in English will require further thought before the suggested two-dimensional characterisation of the language can be confirmed. Furthermore it has not escaped our attention that the definition of the foot used for the two languages is different: the 'full vowel' foot used for English would simply not work for a language such as Estonian where virtually every vowel is 'full'. However, acknowledging that the relevant definition of the foot may be language specific may be no worse than having to acknowledge that, at another level of the prosodic hierarchy, the relevant definition of the syllable has to differ.

Although further conceptual and empirical work will be needed to resolve such complexities, we are confident that orthogonal syllable and foot PVIs will prove effective in capturing the rhythmic properties of at least a subset of languages.

5. Conclusions

This paper presented two experiments addressing the issue of quantitative measures of acoustic correlates of speech rhythm. The first experiment compared a number of PVI measures for Estonian calculated on the basis of various phonological units. It appeared that the PVI measure for the foot was significantly smaller (and consequently a better reflection of Estonian rhythmicality) than other measures such as the syllable and vocalic ones. The second experiment compared the Estonian syllable and foot PVI measures to those of English, showing that foot the foot PVI values in the two languages are comparable whereas the mean PVI for the syllable is significantly higher in English than in Estonian.

It was argued that the now common practice of characterising rhythm in terms of syllable components (vocalic and consonantal intervals) can be improved on by the use of a two-dimensional characterisation using syllable and foot PVIs. This combination gives a more appropriate and subtle account of the rhythm of languages, and in the case of Estonian, examined here in detail, explains the intuition that it is both stress and syllable timed.

6. Acknowledgments

7. References

[1] Abercrombie, D., 1967. *Elements of General Phonetics.* Edinburgh: Edinburgh University Press.

[2] Asu, E.L., 2004. *The Phonetics and Phonology of Estonian Intonation.* Doctoral dissertation. University of Cambridge.

[3] Asu, E.L.; Nolan, F., 2005. Estonian rhythm and the Pairwise Variability Index. *Proceedings, FONETIK 2005.* Department of Linguistics, Göteborg University, 29-32.

[4] Barry, W.J.; Andreeva, B.; Russo, M.; Dimitrova, S.; Kostadinova, T., 2003. Do rhythm measures tell us anything about language type? *Proceedings of the 15th ICPhS.* Barcelona, 2693-2696.

[5] Cruttenden, A., 1997. *Intonation.* Cambridge: Cambridge University Press.

[6] Cutler, A.; Norris D., 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14, 113-121.

[7] Eek, A.; Help, T., 1987. The interrelationship between phonological and phonetic sound changes: a great rhythm shift of Old Estonian. *Proceedings of the 11th ICPhS.* Tallinn, Estonia, 6, 218-233.

[8] Grabe, E.; Low, E.L., 2002. Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven; N. Warner (eds) *Laboratory Phonology* 7, 515-546. Berlin: Mouton de Gruyter.

[9] Kozhevnikov, V.A.; Chistovich, L.A., 1965. *Speech: Articulation and Perception.* Translation: Joint Publications Research Service, US Department of Commerce, 30-543.

[10] Low, E.L., 1998. *Prosodic Prominence in Singapore English.* Doctoral dissertation. University of Cambridge.

[11] Low, E.L.; Grabe, E.; Nolan, F., 2000. Quantitative characterisations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43, 377-401.

[12] Pellegrino, F.; Chauchat, J.-H.; Rakotomalala, R.; Farinas, J., 2002. Can automatically extracted rhythmic units discriminate among languages? *Proceedings of Speech Prosody 2002.* Aix-en-Provence, 563-566.

[13] Ramus, F., 2002. Acoustic correlates of linguistic rhythm: perspectives. *Proceedings of Speech Prosody 2002.* Aix-en-Provence, 115-120.

[14] Ramus, F.; Nespor, M.; Mehler, J., 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 72, 1-28.

[15] Ross, J.; Lehiste, I., 2001. *The Temporal Structure of Estonian Runic Songs.* Berlin: Mouton de Gruyter.

[16] Stockmal, V.; Markus, D.; Bond, D., 2005. Measures of native and non-native rhythm in a quantity language. *Language and Speech* 48, 55-63.

[17] Taylor, D.S., 1981. Non-native speakers and the rhythm of English. *International Review of Applied Linguistics in Language Teaching* 19 (3).

[18] http://www.phon.ox.ac.uk/~esther/ivyweb/