Measuring Basic Tempo across Languages and some Implications for Speech Rhythm

Gertraud Fenk-Oczlon[1], *August Fenk*[2]

[1] Department of Linguistics and Computational Linguistics, University of Klagenfurt, Austria
[2] Department of Media and Communication Studies, University of Klagenfurt, Austria

gertraud.fenk@uni-klu.ac.at, august.fenk@uni-klu.ac.at

Abstract

Basic language-inherent tempo cannot be isolated by the current metrics of speech rhythm. Here we propose the number of syllables per intonation unit as an appropriate measure, also for large-scale comparisons between languages. Applying it to an extended sample of in the meantime 51 languages has not only corroborated our previously reported negative cross-linguistic correlation of this metric with syllable complexity, but has revealed, moreover, significant correlations with several in part directly time-dependent rhythm measures proposed by other authors. We discuss relations between intrinsic tempo and (a) other facets of rhythm and (b) rhythm classifications of language.
Index Terms: tempo, metrics, rhythm, syllable complexity, intonation unit, cross-linguistic correlations

1. Introduction

In view of the common impression that "some languages are spoken more quickly than others", three possibilities are considered in [1]: Some languages really are spoken more rapidly (i); we get this impression because of some sort of illusion (ii); rapid vs. slow speaking is a matter of acceptance in different societies (iii). But why not consider for instance a situation-dependent, intra-individual variability? Thus, the above list of possibilities need not be viewed as exhaustive. And the possibilities need not be viewed as mutually exclusive. We tend to assume both, a variation of tempo *within* and *between* systems at any level, at the level of individual subjects, of different societies, and of different languages. Here we will focus on differences *between languages*.

In the following section we will illustrate some relations between rhythm and tempo. After a short discussion of the methodological problems and possibilities of quantitative comparisons across languages in Section 3, a simple measure for language-specific differences in basic tempo is suggested in Section 4. The advantage of this metric is that it captures the language-inherent or intrinsic tempo of language that is out of reach of most of the metrics (Section 3.1.) developed for measurements of rhythm and speech rate (I); that it is directly based on differences in the linguistic structure of languages (II); that it is quite independent of intra- and inter-individual differences (III) and appropriate for large-scale comparisons (IV). Thus its application also contributes to the general question in [2] whether, or to which extent, "speech tempo is part of the linguistic structure of languages". In Section 5 we present the results of an application of this measure in a sample of 51 typologically different languages. These results are related to the classification of languages as syllable-timed vs. stress-timed and to the respective data reported in [3] and especially in [4]. Implications for a systemic typology are discussed in Section 6.

2. Rhythm and tempo

The focus of our study is on cross-linguistic differences regarding the basic tempo within clauses. But rhythm, as we know, cannot be reduced to the mean size or duration of segments and the mean tempo within them. The basic size of segments and their elements will interact with "local" variations. Such local variations, and other facets of rhythm such as prominence [5, 6], are not, or not directly, captured by our method (Section 4).

Rhythm and tempo are, however, closely related concepts. Fraisse mentions two different relationships that seem to have anticipated or inspired contemporary operationalizations:

"One of the perceptual aspects of rhythmic organization is tempo. It can be lively or slow. It corresponds to the number of perceived elements per unit time, or to the absolute duration of the different values of the duration" [7].

3. Two groups of measures for a comparison of basic tempo across languages

One may distinguish between two groups of relevant metrics: those using chronometry (3.1) and those doing without (3.2).

3.1. Measuring durations

A straightforward determination of tempo as the "number of elements per unit of time" depends on chronometry, and so does any determination of distances or intervals (e.g. the mean distance between vowels) as duration.

In face of the enormous inter-individual differences in speech rate one would need, from all the languages to be compared, a large and representative sample of speakers: "representative" with respect to variables such as dialectal region, age and sex, education and profession. And in face of the enormous intra-individual variation in speech rate we would need a large set of data from every subject in all our samples of speakers. Since hardly any researcher will be able to manage this, the domain of such time-consuming measurements of duration will be restricted to settings investigating speech rate in a rather small number of languages, e.g. as a variable depending on conversational contexts [8].

3.2. Counting procedures

Distances or intervals, such as the mean distance between vowels, can also be expressed in terms of the number of intervening elements, e.g. consonants. The results of such a measurement will of course correspond with the proportion of vowels in a spoken or written text, and this proportion is again based on the results of counting procedures. And the metric "number of elements per *unit time*" must, in such a simple

4 – 4.99		5 – 5.99		6 – 6.99		7 – 7.99		8 – 8.99		9 – 9.99		10 – 10.99	
		Dutch	5.05										
		Czech	5.36										
		Mandarin	5.46										
		Sloven.	5.50	Canton.	6.05								
		German	5.50	Cham	6.32								
		Iceland.	5.59	Bambara	6.46								
		French	5.64	Turkish	6.46								
		Estonian	5.68	Albanian	6.55								
		Russian	5.68	Yoruba	6.59								
		Croatian	5.77	Portug.	6.64								
		English	5.77	Persian	6.64								
		Ewondo	5.77	Finnish	6.73	Hopi	7.12	Kadazan	8.00	Kirundi	9.05		
		Hungar.	5.91	Fukien	6.77	Navaho	7.41	Korean	8.18	Chiquit.	9.14	Hawaiian	10.05
Thai	4.64	Hebrew	5.96	Hindi	6.77	Italian	7.50	Annang	8.23	Javanese	9.18	Roviana	10.05
Vietnam.	4.91	Arabic	5.96	Panjabi	6.77	Greek	7.55	Basque	8.27	Indones.	9.46	Japanese	10.23
Lamso	4.96	Polish	5.96	Macedon.	6.96	Spanish	7.96	Tagalog	8.95	Marranju	9.86	Telugu	10.96

Figure 1: *The frequency distribution of 51 languages over seven classes of the parameter mean n of syllables per clause.*

counting procedure, be replaced by the metric "number of elements per *higher order unit*", as is the case in the number of pulses per musical phrase or per linguistic phrase.

4. Method

In fact, tempo as described in Fraisse is in the first place the number of elements per unit time. The "number of words per minute" would be a corresponding measure, and at a first glance attractive because of the words' role as semantic units. But such a measure not only suffers from the problems of inter- and intra-individual differences in speech rate (3.1.) but also from the fact that the words are enormously variable in their size as measured in syllables. This size reaches from monosyllabic words to whole sentences in polysynthetic languages. We refer to the mean number of elements per unit of time as the "basic tempo", but suggest a metric that does without any direct time measurements.

4.1. The metric: Which units of which higher order unit?

In order to count units within a higher order unit, we need first of all a higher order unit that is both relevant for rhythmic organization and universally comparable. The clause or simple declarative sentence meets these conditions: It corresponds to a prototypical intonation unit that can be characterized as "a sequence of words combined under a single, coherent intonation contour" between two pauses, and a length of typically two seconds [9]. "Intonation units" or "breath groups" [8] are organized in ways not to exceed the duration of a normal exhalation, i.e., a duration that is a culturally relatively independent quantity.

If we consider tempo, the relevant element of this quantity, or the single pulse within this quantity is not the word but the syllable. Thus, in our metric "pulses per higher order unit", the pulses are the syllables and the higher order unit is the intonation unit.

4.2. Procedure

The application of our measure "number of syllables per intonation unit" is simple but requires a controlled set of cross-linguistically comparable and rhythmically relevant units as described above. Native speakers of 51 languages from all continents (19 Indo-European, 32 Non-Indo-European) were asked to translate 22 clauses of this sort. The written translations, or their transcriptions, enabled a counting of the number of words per clause. Furthermore, the subjects were instructed to read their translations in normal speech and to count the number of syllables (which is, apart from determining the borders of the syllables, no problem for the informants). The number of phonemes was determined by the first author, assisted by the native speakers and by grammars of the respective languages.

4.3. Subjects

Most of our 51 informants were students, many of them linguists we met at international conferences. The basic requirement was a good knowledge of either English or German in order to be able to translate the test sentences into their mother tongue. And where necessary they should be able to produce a transcription of their translation.

4.4. Materials

The clauses used were of a special quality: simple declarative sentences encoding one proposition in one intonation unit. Such simple declarative sentences seem to be universal also from a syntactic perspective. Examples for the test sentences using a rather basic vocabulary are: *The sun is shining.* (*Die Sonne scheint.*) *I thank the teacher.* (*Ich danke dem Lehrer.*) *The spring is on the right.* (*Die Quelle ist rechts.*) *Grandfather is sleeping.* (*Der Grossvater schläft.*) *My father is a fisherman.* (*Der Vater ist Fischer.*) A complete list of the sentences in English and in German is presented in [10].

5. Results

Our sample of 51 languages shows a considerable and obviously systematic variation of the number of syllables per simple declarative sentence (Section 5.1). Co-variations with in part directly time-dependent metrics are reported in Section 5.2.

5.1. The more syllables per clause, the fewer phonemes per syllable

Figure 1, which represents the results concerning the number of syllables per sentence, is a frequency distribution

construed according to the principles of exploratory data analysis [11]. It indicates that the location of a certain language on the x-axis (mean number of syllables per simple declarative sentence) is mainly a matter of its mean syllable complexity: Dutch, for instance, is well known for its high mean syllable complexity, and Japanese for its low mean syllable complexity.

In our sample, the mean number of syllables per clause is 7.02, and the mean number of phonemes per syllable is 2.24. German shows the highest mean syllable complexity (2.79 phonemes per syllable), followed by Dutch (2.78) and Thai (2.75). The languages with the lowest syllable complexity – Hawaiian (1.76 phonemes per syllable), Japanese (1.88) and Roviana (1.92) – are localized at the other end of the distribution in our Figure 1.

The results of a cross-linguistic correlation between the syllable complexity and the number of syllables per sentence give: $r = -0.73$ (sign. $p < 0.01$).

This correlation indicates time limits being effective on clause length, forcing a trade-off between the length of syllables in number of phonemes and the length of clauses in number of syllables. It has proved to be very robust. From the first computation in [12] in a sample of 26 predominantly Indo-European languages up to now 51 predominantly Non-Indo-European languages it has shown only little variation within a range from -0.77 to -0.73.

5.2. Correlations with other metrics of rhythm

Can our results be related to other durational metrics mentioned in the literature, according to which languages are classified as stress-timed, syllable-timed, or mora-timed? Is there a relationship between the number of syllables per clause and the rhythm classes stress-timed versus syllable timed?

In a previous study [13] we correlated both our metrics, syllables per clause and phonemes per syllable, with the "normalized Pairwise Variability Index (nPVI)" used in Grabe and Low [3] and found significant coefficients in both cases – despite the fact that in [3] only one speaker from each language was recorded. Partial correlations revealed syllable complexity as the factor that explains most of the variation. Due to these findings we suggested a re-interpretation of the classification of languages as stress-timed, syllable-timed and mora-timed. These are not distinctive classes but at best types that can be assigned more or less loosely to our tempo-related metrics "syllable complexity" and "number of syllables per clause".

In the meantime, with our now extended sample of languages, our metrics can also be related to those suggested in [4] as the most appropriate ones for distinguishing between stress-timed and syllable-timed languages, i.e., the percentage of vowels (%V) and the standard deviation of the duration of consonantal intervals within each sentence (ΔC). Since these metrics reflect properties of syllable structure [14], one has to expect correlations between them and our complexity measures (a and b in Table 1). The list of languages in this table corresponds with the languages of Table 1 in [4]; only Catalan, which is not included in our sample, had to be omitted. We expected high correlations for two reasons:

Since the core of (almost) any syllable is a vowel, a high number of syllables per clause will coincide with a high percentage of vowels (%V). And a high standard deviation of consonantal intervals (ΔC) requires a relatively high number of consonants per syllable. In view of our negative correlation between syllable complexity and the number of syllables per clause the reverse relations had to be expected

in the correlations of the number of phonemes per syllable with %V and with ΔC.

The results are thus: The correlations obtained between our tempo measures and the measures used in [4] turned out to be significant and showed the expected signs (Table 2).

Table 1. *A comparison of our complexity measures (a, b) with two measures suggested in [3]: the proportion of vocalic intervals (%V) and the standard deviation of the duration of consonantal intervals within each sentence (ΔC).*

Languages	a syll/clause	b phon/syll	c %V	d ΔC
English	5.77	2.58	40.1	5.35
Polish	5.95	2.43	41.0	5.14
Dutch	5.05	2.78	42.3	5.33
French	5.64	2.16	43.6	4.39
Spanish	7.96	2.09	43.8	4.74
Italian	7.50	2.12	45.2	4.81
Japanese	10.23	1.88	53.1	3.56

Table 2. *Some correlations between the four columns in Table 1.*

	c % V	d ΔC
a syll/clause	$r_{ac} = 0.89$ $p < 0.01$	$r_{ad} = -0.81$ $p < 0.05$
b phon/syll.	$r_{bc} = -0.76$ $p < 0.05$	$r_{bd} = 0.87$ $p < 0.05$

6. Discussion

In order to summarize: The 51 languages in our sample show a considerable variation in the number of syllables per intonation unit, ranging from 4.64 in Thai up to 10.96 in Telugu. Our previously found cross-linguistic negative correlation between this variable and the number of phonemes per syllable was corroborated in this extended sample of languages. We view these differences as differences in the intrinsic tempo of languages. The higher the number of syllables per intonation unit, the smaller the syllables and the higher the tempo – similar to music where in "phrases containing many notes, the notes are usually very fast" [15]. Significant correlations between our metrics with well-known metrics differentiating between stress-timed and syllable-timed languages indicate that languages traditionally classified as stress-timed tend to have a slower tempo than syllable-timed languages. Languages usually classified as mora-timed, such as Japanese, Telugu and Hawaiian, exhibit the highest number of syllables per intonation unit. Due to their simple syllable structure they need a higher number of syllables to encode a certain semantic unit, i.e., to form a certain proposition. They produce more pulses per time and per higher order unit, i.e., per clause or intonation unit. Languages characterized by a rather small number of syllables per proposition produce a lower number of pulses per intonation unit but need a higher number of phonemes per syllable. Thus, a higher intrinsic tempo of a certain language does not indicate a higher information transmission rate in that language. Only within the very same language will an increase of speech rate increase that transmission rate.

Syllable structure not only influences the number of syllables per clause or the basic tempo of languages. In a

previous study [16] we found significant cross-linguistic correlations between syllable complexity and other metric and non-metric properties of language: Languages with complex syllable structure tend to have a high number of words per clause, a low number of syllables per word and VO word order, whereas languages with simple syllable structure tend to have a low number of words per clause, a high number of syllables per word and OV word order. And higher syllable complexity is also associated with a larger phonemic inventory and with a higher proportion of monosyllabic words [17].

How can these interactions between syllable complexity and other properties of language be related to speech rhythm? In [5] it is argued that prominence or stress is a crucial concept of rhythm and that "one difference between languages called stress-timed and those called syllable-timed may have to do with the spacing of prominences, not in terms of duration, but in terms of number of syllables". Inter-stress intervals in terms of number of syllables are longer in syllable-timed languages than in stress-timed languages. We believe that word length is an important factor that might influence the spacing of prominences. Monosyllabic content words are usually stressed [18]. Thus, languages with a high proportion of monosyllabic words will tend to shorter inter-stress intervals. Indeed, the languages in our sample traditionally classified as stress-timed tend to have a high proportion of monosyllabic words.

7. Conclusions

We suggested and applied the simple metric "syllables per intonation unit" in order to analyze and compare languages with respect to their intrinsic tempo. Furthermore, significant correlations were found between our metric "number of syllables per intonation unit" and other metrics proposed in the literature for differentiating between languages labeled as stress-timed vs. syllable-timed. Languages traditionally classified as stress timed tend to have a lower intrinsic tempo than languages classified as syllable-timed or mora-timed. Likewise, our previous cross-linguistic correlations between syllable structure and word length suggest interactions between prominence and inter-stress intervals on the one hand and intrinsic tempo of languages on the other.

The whole body of findings amounts to the assumption that the key to an understanding of the perceived differences between the languages' rhythm, including tempo, lies in their different syllable structure. Languages showing a high tempo according to our measure tend to have simpler syllable structures. Therefore, their higher intrinsic tempo does not at all imply a faster transmission of propositions.

8. Acknowledgements

9. References

[1] Roach, P., "Some languages are spoken more quickly than others", in L. Bauer and P. Trudgill [Eds], Language Myths, 150-158, Penguin, 1998.

[2] Verhoeven, J., De Pauw, G. and Kloots, H., "Speech rate in a pluricentric language: A comparison between Dutch in Belgium and in the Netherlands", Language and Speech, 47(3):297-308, 2004.

[3] Grabe, E. and Low, E. L., "Durational variability in speech and the rhythm class hypothesis", in C. Gussenhoven and N. Warner [Eds], Laboratory Phonology, 7:515-546, Mouton de Gruyter, 2002.

[4] Ramus, F., Nespor, M. and Mehler, J., "Correlates of linguistic rhythm in the speech signal", Cognition, 73:265-292, 1999.

[5] Arvaniti, A., "Rhythm, timing and the timing of rhythm", Phonetica, 66(1-2): 46-53, 2009.

[6] Kohler, K., "Rhythm in speech and language. A new research paradigm", Phonetica, 66(1-2):29-45, 2009.

[7] Fraisse, Paul., "Rhythm and Tempo", in D. Deutsch [Ed], The Psychology of Music, 149-180, Academic Press, 1982.

[8] Yuan, J., Liberman, M. and Cieri, C., "Towards an integrated understanding of speaking rate in conversation", Interspeech 2006.
Online:http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.8110, accessed on 17 Feb 2009.

[9] Chafe, W., "Cognitive constraints on information flow", in R.S. Tomlin [Ed], Coherence and Grounding in Discourse, 21-25, John Benjamins, 1987.

[10] Fenk-Oczlon, G. and Fenk, A., "Cognition, quantitative linguistics, and systemic typology", Linguistic Typology, 3: 151-177, 1999.

[11] Tukey, J., "Exploratory data analysis", Addison-Wesley, 1977.

[12] Fenk-Oczlon, G. and Fenk, A., "The mean length of propositions is 7 plus minus 2 syllables – but the position of languages within this range is not accidental", in G. d'Ydevalle [Ed], Cognition, Information Processing, and Motivation, 355-359, North Holland, 1985.

[13] Fenk-Oczlon, G. and Fenk, A., "Some parallels between language and music from a cognitive and evolutionary perspective", Musicae Scientiae, 201-226, Special Issue, 2009-2010.

[14] Pfitzinger, H. R., "Phonetische Analyse der Sprechgeschwindigkeit", Forschungsberichte (FIPKM) 38, 117-264, Inst. Für Phonetik und Sprachliche Kommunikation der Univ. München, 2001.

[15] Temperley, D., "The cognition of basic musical structures", MIT Press, 2001.

[16] Fenk, A. and Fenk-Oczlon, G., "Crosslinguistic computation and a rhythm-based classification of languages", in M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nürnberger and W. Gaul [Eds], From Data and Information Analysis to Knowledge Engineering, 350-357, Springer, 2006.

[17] Fenk-Oczlon, G. and Fenk, A., "Complexity trade-offs between the subsystems of language", in M. Miestamo, K. Sinnemäki and F. Karlsson [Eds], Language Complexity: Typology, Contact, Change, 43-65, John Benjamins, 2008.

[18] Pike, K., "The Intonation of American English", University of Michigan Press, 1945.