Workshop on Typology of African Prosodic Systems (TAPS)
Universität Bielefeld, 18−21 May 2001

Rhythm in African tone languages: a study of Ibibio, Anyi and Ega

(Long abstract)

Ulrike Gut, Eno−Abasi Urua, Sandrine Adouakou, Dafydd Gibbon
Universität Bielefeld, Germany
{gut,urua,adouakou,gibbon}@spectrum.uni−bielefeld.de

Speech rhythm is a much discussed issue in phonetics and phonology. Impressionistic accounts agree that the languages of the world differ in their rhythm (syllable−timing, stress−timing, mora−timing) but attempts to capture these differences acoustically have so far been unsatisfactory. The aim of this paper is to describe three West African tone languages − Anyi, Ega and Ibibio − which have not been previously classified in terms of their speech rhythm.

The rhythm of the languages of the world have traditionally been divided into stress−timed and syllable−timed (Pike 1945, Abercrombie 1967). Rhythm is understood to be a periodic recurrence of events, which in stress−timed languages are stress beats and in syllable−timed languages syllables. Abercrombie (1967) sees speech rhythm as "essentially a muscular rhythm, and the muscles concerned are the breathing muscles" (p. 96). In syllable−timed languages "chest−pulses, and hence the syllables, recur at equal intervals of time − they are isochronous" (p. 97). Syllables are assumed to be equal in length (p. 98), stress−pulses, on the other hand, are unevenly spaced. Abercrombie cites Yoruba, a West African tone language, as an example for a syllable−timed language. Stress−timed languages such as English, in contrast, are supposed to have regular recurring stress beats with the time interval separating two beats of equal length. Since the number of syllables between two stress beats varies, their length is adjusted to fit into the stress interval − syllable length, hence is very variable in stress−timed languages.

Acoustic correlates of these claims have not been found. Neither are interstress intervals in stress−timed languages of equal length (Classé, 1939; Bolinger, 1965; O'Connor, 1965; Uldall, 1971; Hill et al., 1979; Faure et al., 1980; Roach, 1982; Dauer, 1983), nor is the variation between syllable length different between languages classified as stress−timed and those classified as syllable−timed (Roach, 1982).

Recently, it has been suggested that speech rhythm reflects variety of syllable structures, phonological vowel length distinctions, absence/presence of vowel reduction and lexical stress (Dauer, 1983). Whereas languages classified as stress−timed show a variety of different syllable structures, languages classified as syllable−timed have a majority of CV syllables. Furthermore, differences in rhythm between languages reflects whether a language has vowel reduction or not where those classified as stress−timed usually do. In addition, syllable−timed languages either do not have lexical stress or accent is realized by variations in pitch contour. Conversely, stress−timed languages realize word level stress by a combination of length, pitch, loudness and quality changes, which result in clearly discernible beats.

With the exception of Roach (1982), the speech rhythm of West African languages has not been investigated yet. In this paper, we selected three of these languages, Anyi, Ega and Ibibio. We started from the first working hypothesis that they are syllable−timed, as reflected in a low variation between syllable durations. Next, we tried to correlate phonological features such as absence/presence of phonemic vowel length, syllable structure, tone system with the speech rhythm. We use the following definition of rhythm, which is related to that of Abercrombie:

> *A rhythmic pattern is a sequence of adjacent intervals perceived as having equal duration, each associated with the same trajectory of values of some parameter.*

On the (not uncontroversial) assumption that perceived equality of duration correlats approximately with measured equality of duration, we defined a simple measure of durational differences with a range between 0 and 100, which we refer to as Rhythm Ratio (RR): for a sequence of units (e.g. syllables) of length, the average ratio of the durations of the syllables is calculated:

$$RR = 100 * \frac{\sum_{k=1}^{m-1} \frac{d_i}{d_j}}{m-2} \quad where \;\; i=k,\; j=k+1 \;\; if \; d_i \leqslant d_j, \;\; else \; j=k, i=k+1$$

The condition ensures that the larger of the two duration values is the denominator of the duration division, and the smaller is the numerator. The larger the RR, the less similar are the durations of adjacent intervals (with respect to the relevant parameter), and therefore the more rhythmical the overall pattern. If the syllable is taken as the interval−defining unit, then a smaller RR is more likely to indicate syllable timing, and a larger RR is more likely to indicate foot (or stress, etc.) timing. The value of RR is 0 for temporally absolutely identical durations and approaches the limit of 100 for extremely dissimilar durations, but empirically plausible values of the RR are likely to be between 10 for very similar intervals and 50 for very dissimilar intervals. There are several underlying assumptions which underlie this procedure, of course, which may be regarded as problematic. However, we are interested in developing a first approximation and therefore do not engage in further discussion of these assumptions at this point.

In order to permit rapid comparisons of different measures, the RR formula was implemented as a program which takes a Transcriber or a Praat annotation file as input and outputs a range of statistics (min, max, range, sum, mean, median, mode, variance, standard deviation, standard error, confidence intervals) about durations, absolute differences between adjacent duration pairs, normalised differences between adjacent duration pairs, as well as z−scores for durations and duration pair differences. We also introduced a binary classifier for predicting typical focal and nonfocal units, which we refer to as the Focal−Nonfocal Measure (FNM). So far we have tested simple classifiers with boundaries based on a number of criteria such as mean syllable length, the mid value between minimum and maximum syllable length, and are planning to use further types.

The corpus consisted of recordings from one Anyi speaker telling a story, one Ega speaker delivering a formal address and one Ibibio speaker also delivering a formal address were analysed using Praat and Transcriber.

We have obtained the following results:

1. The RR for adjacent syllable durations in Anyi, Ega and Ibibio are very similar (less than 20), while for English RR is considerably higher (more than 25).
2. Syllables classified as short and long by the FNM were more similar in length for Anyi, Ega and Ibibio than for English.

We argue that this is due to the difference in syllable structure, absence of lexical stress and tone system in the West African tone languages as compared to British English, that these differences lead to a closer approximation to syllable timing (syllable isochrony) in Anyi, Ega and Ibibio than in English, and that our measure plausibly contradicts previous claims that the syllable−foot timing distinction is not valid.