Indian Speech and Hearing Association

(Regd. under the Karnataka Societies Registration Act. Karnataka Act No. 17) Registration No. S 25/67-68

Dr. N. Rathna Oration

41st ISHACON Pune 2009

Speech Rhythm in Indian Languages

Dr. S. R. Savithri, Professor, Department of Speech-Language Sciences
All India Institute of Speech & Hearing, Manasagangothri, Mysore 570 006

At the outset I would like to thank the ISHA organizing committee for giving me the honour of Dr. N. Rathna oration. Dr. Rathna was my teacher and guide of the doctoral studies. Today I will speak on ***speech rhythm*** in Indian languages which is a neglected area of research.

Word recognition is the heart of speech recognition; but it requires listeners to turn a continuous stream of speech into a percept consisting of discrete, non-overlapping individual lexical units. This process of segmentation works wonderfully well in the native language, but not always so well for languages learned later in life. One reason for this is that segmentation procedures are in part language-specific. There are in fact many ways in which listeners tailor their speech processing, and in particular the processes by which they segment continuous speech, such that great efficiency is achieved with native-language input; any aspect of phonological structure can be exploited in this way. Language rhythm is an aspect of phonology which definitely plays a role in listening.

That languages differ in characteristic rhythm has long served as an inspiration in phonetic research (Pike, 1945; Abercrombie, 1964), even though no consensus has emerged on how the undoubted differences in rhythmic structure should be captured (Cutler, 1991). A new type of approach in which rhythmic structure is expressed via a formulation of segmental and syllabic patterning (Ramus, Nespor & Mehler, 1999; Low, Grabe & Nolan, 2000) offers perspectives of innovative advance in this area. Moreover, this approach offers at last the possibility of a more solid foundation for the notion of rhythmic classes into which languages may be organized. This notion can be traced to Abercrombie (1964) who distinguished stress-timed from syllable-timed languages. The class of mora-timed languages was later added (Hoequist, 1983).

Rhythm refers to an event repeated regularly over a period of time. Three types of rhythm have been identified in the past. A language which has very simple syllabic structure (for e.g. V, CV, CCV) will have almost equal syllabic duration or the difference between the simplest and the most complicated syllable is not wide, and it is possible to say that any syllable is less than 330 ms. This is termed mora-timed rhythm. If a language has a little more complicated type of syllable structure (for e.g. CCV, CVC, CCCV) then the syllable duration is not so equal as that in a mora-timed rhythm and it is termed syllable-timed language. In a syllable timed language, the difference between the successive intervocalic duration is high and that between successive vocalic duration is low. But in mora-timed language the difference between the successive intervocalic and vocalic duration is low. If a language has complex syllables such as ones with consonant clusters, the difference between syllables can be very wide, such as 'a' (around 150 ms) and strength (around 620 ms) in English. In this case there is a slow stress-timed syllable.

Rhythm Class Hypothesis (henceforth RCH) states that each language belongs to one of the prototypical rhythm classes known as either syllable-timed, stress-timed or mora-timed. In an earliest attempt Abercrombie (1967) measured the ***syllable duration*** in French, Telugu, Yoruba, English, Russian, and Arabic which were classified as belonging to different rhythm types. He reported that the syllable durations were almost equal across languages thus not supporting the RCH. Following this, Roach (1982), for instance, compared ***interstress intervals*** in languages classified as stress-timed and syllable-timed. He investigated two claims made by Abercrombie (1967) about the difference between stress-timed and syllable-timed rhythm (a) there is considerable variation in syllable length in a language spoken with stress-timed rhythm, whereas in a language spoken with syllable-timed rhythm, syllables tend to be equal in length, and (b) in syllable-timed languages, inter-stress intervals are unevenly spaced. Roach's (1982) findings did not support either claim. The syllable-timed languages in his sample exhibited greater variability in syllable durations than the stress-timed languages. Roach also observed a wider range of percent deviations in inter-stress intervals in stress-timed than in syllable-timed languages. Roach concluded that measurements of ***time intervals*** in speech could not provide evidence for rhythm classes.

Ramus, Nespor & Mehler (1999) measured ***vowel durations*** and the ***duration of intervals between vowels*** in a set of tightly controlled sentences from eight languages (5 sentences each produced by four speakers = 160 sentences). Ramus and colleagues argued that a viable account of speech rhythm should not rely on complex and language-dependent phonological concepts but on purely phonetic characteristics of the speech signal. These authors segmented speech into vocalic and consonantal intervals. Ramus et al. computed three acoustic correlates of rhythm from the measurements: (a) ***%V***, the proportion of time devoted to vocalic intervals in the sentence, disregarding word boundaries; (b)***ΔV*** : ***the standard deviation of vocalic intervals; (c) ΔC: the standard deviation of consonantal intervals, the sections between vowel offset and vowel onset***. On the basis of their findings, they reported that a combination of ***%V and ΔC*** provided the best acoustic correlate of rhythm classes. In English, which has full and reduced vowels, %V was smaller than in French, which does not have vowel reduction. On the other hand, ΔC was larger in English and reflected the more complex syllable options available in that language. The results found that along two dimensions (%V and ΔC), languages are not scattered randomly, but are clustered in groups that strongly resemble rhythm classes: English, Dutch and Polish as stress-timed languages, French, Spanish, Italian and Catalan as syllable-timed languages, and Japanese as a mora-timed language. Figure 1 recalls the main results and figure 2 shows the position of languages as per %V and △C.

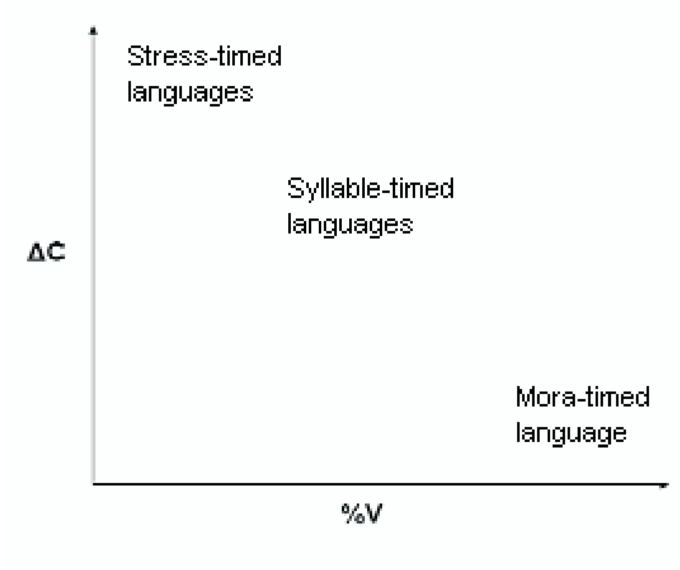

Figure 1: Illustration of stressed and syllable timed languages (Ramus et. al., 1999).

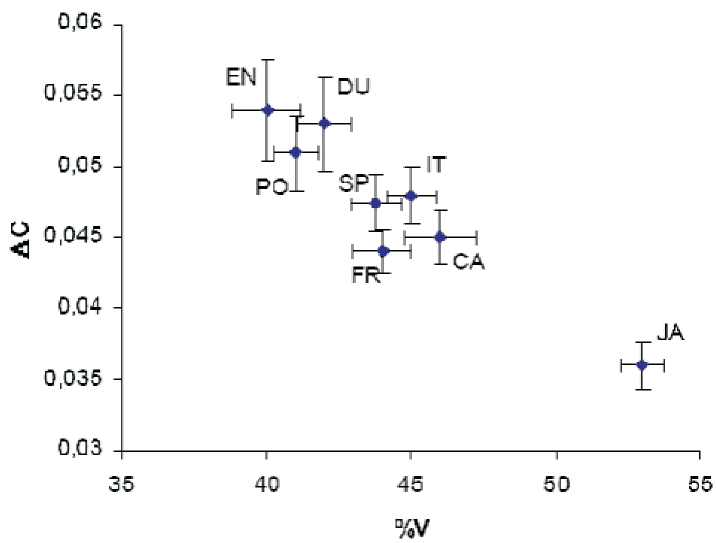

Figure 2: Standard deviation of consonantal intervals vs. proportion of vocalic intervals (EN- English, DU- Dutch, PO- Polish, FR- French, SP- Spanish, IT- Italian, CA- Catalan, JA- Japanese).

The results indicate that the measures taken reflect rhythmic differences, but not classes. It is indeed entirely possible that when more languages are added, the clusters will be drowned in a uniform rhythmic continuum or space.

The ***Pairwise Variability Index*** (PVI) is a quantitative measure of acoustic correlates of speech rhythm, which calculates the pattering of ***successive vocalic and intervocalic (or consonantal) intervals*** showing how one linguistic unit differs from its neighbour (Low, 1998). Grabe & Low (2000) developed "normalized Pairwise Variability Index" (nPVI) for rhythmic analysis of vocalic durations. The raw Pairwise Variability Index" (rPVI) is used for rhythmic analysis of intervocalic durations. In an independent study, Ramus (2002) examined the duration and variability of vocalic and intervocalic intervals in eight languages. The rationale behind the consideration of inter-vocalic intervals is that stress-timed languages also tend to allow more complex syllables, and therefore longer and more variable sequences of consonants than syllable-timed languages. Duration of vowels, and the duration of intervals between vowels (excluding pauses) are measured. Then a Pairwise Variability Index is computed for each type of measurement. The index expresses the level of variability in successive measurements. The raw Pairwise Variability Index (rPVI) is given in equation (1).

$$rPVI = \frac{100}{m-1} \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right] \quad \textbf{Equation 1}$$

Where, m is number of intervals, vocalic or intervocalic, in the text and d is the duration of the k^{th} interval. Notice that rPVI is not normalized for speech rate.

Low et al. used a normalized version of the Pairwise Variability Index in their measurements on vowel durations. The equation for this version, the normalized Pairwise Variability Index (nPVI), is as follows :

$$nPVI = \frac{100}{m-1} \times \left[\sum_{k=1}^{m-1} \left| d_k - d_{k+1} \right| / (m-1) \right] \quad \textbf{Equation 2}$$

Where, m is number of items in an utterance and d is the duration of the k^{th} item.

Equation (2) shows that the nPVI is compiled by calculating the difference in duration between each pair of successive measurements, taking the absolute value of the difference and dividing it by the mean duration of the pair. Equation (1) for the rPVI differs only in omitting the third step. The differences are then summed and divided by the number of differences. The output is multiplied by 100, because the normalization produces fractional values.

This approach has successfully shown an empirical difference between so-called stress-timed languages and syllable-timed languages, with less contrastiveness of successive vocalic durations for syllable-timed languages. Similarly, several studies (Warner & Aria, 2000; Ramus, 2002) revealed distinct rhythmic features of Japanese, which belongs to mora-timed languages. Table 1 summarizes the basic characteristics of each language class regarding relative values of vocalic nPVI and intervocalic rPVI.

Language class	Languages	Intervocalic rPVI	Vocalic nPVI
Stress-timed	English, Germany	High	High
Syllable-timed	French, Spanish	High	Low
Mora-timed	Japanese	Low	Low

Table 1: Summary of basic characteristics of each language class regarding relative values of vocalic nPVI and intervocalic rPVI.

The present report is based on an investigation of rhythm in 12 Indian languages Assamese, Bengali, Gujarathi, Hindi, Kashmiri, Marathi, Oriya, Punjabi, Kannada, Malayalam, Tamil and Telugu (Dravidian languages).

A 100-word passage as read by 20 normal speakers (10 F, 10 M) in each language was audio-recorded and displayed as a waveform. The PVIs were calculated. Two way ANOVA indicated significant difference between languages on nPVI (vocalic) values [F (11, 216) = 10.377, (p<0.01)]. There was a significant difference between gender [F (1,216) = 4.091, (p<0.05)] and interaction between language and gender [F (11, 216) = 2.005, (p<0.05)]. Hence Duncan's post hoc test was done for all the languages.

The rPVI (intervocalic) values also showed significant difference between languages [F (11,216) = 4.046, (p<0.01)] and interaction between language and gender [F (11,216) = 2.439, (p<0.05)] respectively. But there was no significant difference between gender [F (1,216) = 0.622, (p>0.05)]. Duncan's post hoc test revealed significant difference between rPVI values among different languages.

The results show that all languages except Hindi were mora-timed language and Hindi was syllable-timed language. Hence the rhythm of the language can be taught on this basis. For example, syllable-timed approach will be appropriate for Hindi and mora-timed for other languages. Figure 3 shows mean of nPVI and rPVI in reading task.

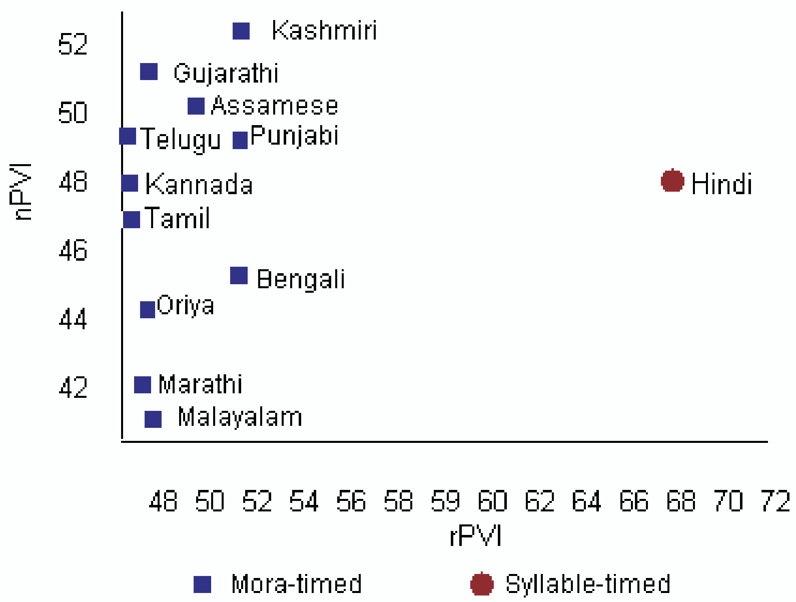

Figure 3: Mean PVIs in reading task

The results have positive implications for a speech pathologist. Rhythm can be taught to patients with aprosodia provided one knows the type of rhythm in a language. For example, a visual feedback of equal syllable timing can be taught to a patient speaking languages other than Hindi. However, as this data is based on reading sample, caution should be exercised for using such procedures in speech. Also, prolongation therapy may suite speakers of all these languages, as they are mora / syllable-timed languages.

We already know that the speech of hearing impaired individuals lacks rhythm. However, till date it is undefined. Now that there is normative data, the rhythm used by hearing impaired individuals can be investigated.

References

Abercrombie, D. (1964). “A phonetician's view of verse structure,” *Linguistics*, Vol. 6, pp. 5-13.

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine Pub. Co.

Cutler, A. (1991). “Linguistic rhythm and speech segmentation,” in *Music, Language, Speech and Brain*, J. Sundberg, L. Nord and R. Carlson, (Eds.), pp. 157-166. London: Macmillan.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.). *Laboratory phonology*, 7, 515-546. Berlin: Mouton de Gruyter.

Hoequist, C.,E. (1983). “Syllable duration in stress-, syllable and mora-timed languages,” *Phonetica*, Vol. 40, pp. 203-237.

Low, E. L. (1998). *Prosodic prominence in Singapore English*. Unpublished Ph.D. Thesis, University of Cambridge.

Low, E. L., Grabe, E. & Nolan, F. (2000). Quantitative characterizations of speech rhythm: 'syllable-timing' in Singapore English. *Language and Speech*, 43, 377-401.

Pike, K., L. (1945). *The Intonation of American English*. Ann Arbor: The University of Michigan Press.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perceptives. *Proceedings of Speech Prosody*, 115-120.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 72, 1-28.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.). *Linguistic Controversies*, pp. 73-79, London: Arnold.

Warner, N., & Arai, T. (2000). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of India*, 109, 1144-1156.