Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

An analysis of the distribution of syllables in prosodic phrases of stress-timed and syllable-timed languages

Jeena J Prakash[1], *Hema A Murthy*[2]

[1,2]Indian Institue of Technology Madras
jeenaj,hema@cse.iitm.ac.in

Abstract

This paper presents an analysis of syllable rhythm in different types of languages, namely syllable-timed and stress-timed. Four Indian languages that are syllable-timed and American English and Scottish English, that are stress-timed are used for the study. This analysis attempts to bring the similarity or differences between the two types of languages in terms of syllable rhythm at the level of prosodic phrases. Two different studies are performed, namely, the number of syllables that make up a prosodic phrase and the ratio of number of syllables in adjacent prosodic phrases. Different probability distributions are fitted to the data and the goodness of fit is determined using quantile-quantile plot. The key contribution of this paper is the observation that the total number of syllables in a sentence, number of syllables in individual prosodic phrases and the ratio of number of syllables between pairs of adjacent prosodic phrases in declarative sentences uniformly (across all languages) follow a Gamma distribution.

Index Terms: prosodic phrases, number of syllables, ratio of number of syllables, stress-timed, syllable-timed, Gamma distribution

1. Introduction

Languages are classified as either stress-timed or syllable-timed based on speech rhythm [1][2][3]. Rhythm is defined either with respect to every syllable in an utterance or with respect to stressed syllables in an utterance. In syllable-timed languages, rhythm is defined as an equal duration between the production of successive syllables in a word. There is an equal prominence associated with every syllable and generally do not result in reduced vowel lengths. Stress-timed languages are those in which there is an equal duration between stressed syllables, and the unstressed syllables are shortened or lengthened accordingly.

Various analyses have been carried out which bring out the similarities and differences between both classes of languages. Rhythm measures and rhythmic classification based on duration of vocalic and consonantal intervals, structure of syllables, duration of vowels and intervals between vowels, presence or absence of vowel reduction, word stress etc., are found in the literature [4][5][6]. Inspite of the studies about stressed and unstressed syllables, duration of syllables and the syllabic structure, study related to the syllabic rhythm at the level of phrases in the two classes of languages is seldom found in literature.

This paper studies the similarities between stress and syllable-timed languages by quantifying rhythm in two ways, namely, (a) number of syllables that make up a phrase (b) ratio of number of syllables in adjacent phrases. The individual prosodic phrases are identified using acoustic cues from the speech waveform during segmentation.

The study carried out in this paper analyses the frequency distribution of phrase length measured in terms of number of syllables that make up the phrase and the ratio of number of syllables in each pair of adjacent prosodic phrases. Different probability distributions such as exponential, lognormal, Gamma and Weibull are fitted to the data. These distributions are special cases of generalized gamma distribution for different choice of its parameters [7]. The goodness of fit of the data to the distribution is determined using the quantile-quantile plots (QQ plots). The QQ plots compares the quantiles of expected distribution with the quantiles of actual data. If the data fits the expected distribution well, then the quantiles of the data lies close to a straight line [8][9].

Indian languages belonging to two different families are studied. Declarative sentences of four Indian languages, namely, Hindi, Bengali, Tamil and Malayalam, which are syllable-timed are analysed; the first two belongs to Aryan and the next two belongs to Dravidian languages [10][11]. This study is later extended to Indian English (English spoken by an Indian) to study the influence of L1 rhythm on L2. Finally, a truly stress-timed language English (spoken by an American and a Scotsman) is studied.

The rest of the paper is organised as follows. Section 2 tells about the preparation of data used for the study. Section 3 explains the data analysis and results. Section 4 concludes the work.

2. Materials and Methods

This section describes various procedures involved in data preparation. Punctuation in text is not commonly used in Indian languages. Phrase boundaries are therefore obtained from the acoustic data, where pauses are used as cues for phrase boundaries.

2.1. Database used

The database consists of speech wave files of four Indian languages, Indian English, American English and Scottish English, recorded in a noise free environment in a studio by professional speakers of the corresponding language. Each language is spoken by a single native speaker of that language. Speakers were asked to read the sentences given to them in a natural way. The sentences used for the recordings were collected from various domains-news, sports, stories etc. American English and Scottish English data is obtained from the CMU database [12]. The details of the data used is given in Table 1.

 doi: 10.21437/SpeechProsody.2016-11

Table 1: *Language databases used.*

Language	No. of hours
Hindi	4
Bengali	2
Tamil	4.3
Malayalam	2
Indian English	5
American English	1
Scottish English	1

2.2. Letter to Sound Rules

Appropriate letter to sound rules are developed for every language to convert the graphemes to syllables. Language specific rule based parsers are used to get the syllabification of words for Indian languages [13]. For syllabifying Indian English, NIST 's tsylb2 syllabification program is used with few corrections based on specific rules [14]. For American and Scottish English, Classification and Regression Tree (CART) based approach is used to get the phonemes from the words [15].

2.3. Segmentation of speech

Automatic segmentation of the speech data is performed at the syllable level. Hybrid segmentation approach in which Hidden Markov Models (HMMs) are used in tandem with group delay is used to obtain accurate syllable boundaries for Indian languages and Indian English [16]. Phrase breaks are obtained using the position of pauses in the segmented speech. The silences in the speech segment which exceeds a threshold of 50 ms is marked as a pause in the segmented speech. The correctness of the position of pauses is verified manually. Using the syllable segmentation algorithm, number of syllables in each phrase are obtained. Segmentation is performed at phone level for American and Scottish English using CMU Sphinx [12]. The number of syllables in each phrase is then obtained by counting the number of vowels and diphthongs between two pauses [17].

3. Data analysis and results

The studies performed in this paper analyse the probability distribution of the number of syllables in individual prosodic phrases and the ratio of number of syllables between adjacent prosodic phrases in declarative sentences. The analysis is performed at the sentence level. Sentences are categorised based on the number of prosodic phrases that make up the sentence. The analysis is performed for each of the categories separately also. The total number of sentences and the total number of prosodic phrases in each language is given in Table 2.

Within each group of sentences considered, the study is again divided into different types. In the first category, the probability distribution of the number of syllables in individual prosodic phrases is analysed for all phrases in the database. In the next study, the distribution of number of syllables in specific prosodic phrases (for example, in the first phrase of all sentences) is analysed. In the third category, probability distribution of the total number of syllables in every sentence is studied.

In the study of the distribution of adjacent phrases, two different studies are again performed, namely, (i) ratio of number of syllables across any pair of adjacent phrases, (ii) ratio of number of syllables between specific adjacent phrases (for example, between 1st and 2nd phrase).

3.1. Analysis of the number of syllables in prosodic phrases

3.1.1. Number of syllables in individual prosodic phrases

The histograms obtained for the number of syllables in individual prosodic phrases suggest that skewed exponential distributions are perhaps appropriate for consideration. Tests of fit are performed with QQ plots using exponential, lognormal, Gamma and Weibull distributions. Figure 1 gives the QQ plots of the number of syllables in individual prosodic phrases for Hindi for all four probability distributions. From the QQ plot it can be seen that the quantiles of the data points of Hindi lies more or less on a straight line for Gamma and Weibull distributions, but fits the Gamma distribution better. Studies on all languages show that Gamma distribution fits better for all languages except Bengali, American English and Scottish English for which a better fit was obtained using the Weibull distribution.

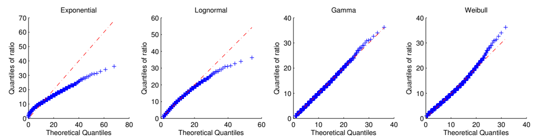

Figure 1: QQ plot of exponential, lognormal, Gamma and Weibull distribution fits for the number of syllables in individual phrases for Hindi

The histogram with distribution of the number of syllables in individual prosodic phrases and the corresponding QQ plots with Gamma distribution is shown in Figures 2 and 3 respectively for all languages considered for the study. From the QQ plots, it is seen that all the data points lie more or less on a straight line, except a small deviation towards the end where the number of syllables is large for Bengali, American English and Scottish English, thus indicating that the number of syllables in individual prosodic phrases indeed follow Gamma distribution. Figure 4 shows the QQ plot of the number of syllables in individual prosodic phrases in Bengali, American English and Scottish English, with Weibull distribution. It is observed that the data fits a Gamma distribution but fits Weibull distribution better for Bengali, American English and Scottish English. The error at 95% confidence for Weibull and Gamma distribution respectively is 0.0029 and 0.0030 for Bengali, 0.0032 and 0.0077 for American English and 0.0041 and 0.0069 for Scottish English. The distributions of the number of syllables in individual phrases for all languages are shown in Figure 5(a).

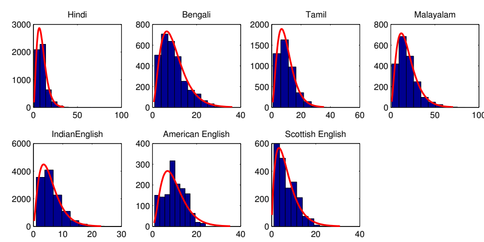

Figure 2: Histogram of the number of syllables in individual prosodic phrases overlaid with the fitted distribution

Table 2: *Total number of sentences and total number of prosodic phrases in different languages.*

Language	Hindi	Bengali	Tamil	Malayalam	Indian English	American English	Scottish English
No. ofsentences	1711	1304	1399	1291	4191	1132	1138
No. of prosodic phrases	5274	2990	4429	2032	11685	1406	2044

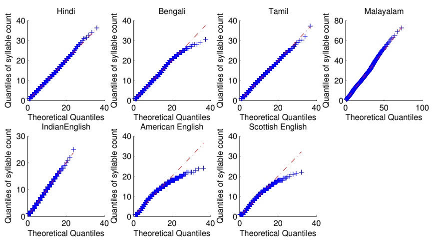

Figure 3: QQ plot of the number of syllables in individual prosodic phrases with Gamma distribution

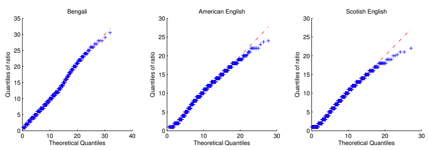

Figure 4: QQ plot of the number of syllables in individual prosodic phrases for Bengali, American English and Scottish English with Weibull distribution

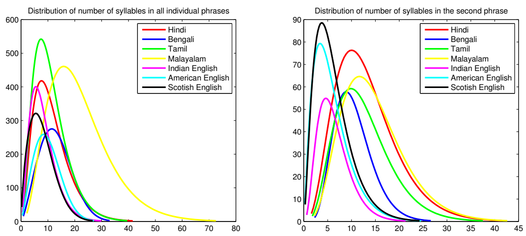

Figure 5: Gamma distribution of the number of syllables (a) All phrases (b) 2nd phrase

3.1.2. Number of syllables in specific prosodic phrases

The study of the number of syllables in specific prosodic phrases showed that they follow Gamma distribution for all languages. The Gamma distribution of the number of syllables in the second phrase of all the sentences is shown in Figure 5(b) for all languages. QQ plot for the same is shown in Figure 6 for all languages. The straight line in the plot indicates that the data follows a Gamma distribution.

The same analysis is performed on different categories of sentences. It is observed that for Bengali and Scottish English, the distribution of number of syllables in the first phrase of single phrase and two phrase sentences, and for American English, the distribution of number of syllables in the first phrase of two phrase sentences fits a Gamma distribution, but fits a Weibull distribution better. For the number of syllables in the second phrase of two phrase sentences, the error at 95% confidence for Gamma and Weibull distribution respectively is 0.0016 and 0.0043 for Bengali, 0.0030 and 0.0052 for American English and 0.0039 and 0.0068 for Scottish English. For the number of syllables in the first phrase of single phrase sentences the error for Gamma and Weibull distribution respectively is 0.0013 and 0.0033 for Bengali and 0.0018 and 0.0027 for Scottish English. For all the other phrases of sentences, and for all other languages the data fits Gamma distribution best.

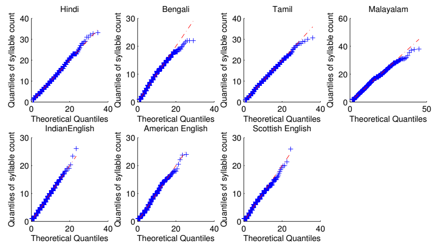

Figure 6: QQ plot of the number of syllables in the second phrase of all sentences with Gamma distribution

3.1.3. Total number of syllables in a sentence

The total number of syllables in a sentence is analysed for all languages and is found that the data follows Gamma distribution for all languages. The Gamma distribution of the total number of syllables in a sentence is shown in Figure 7 for all languages. The same set of sentences are used for American and Scottish English and hence the distribution overlaps. The shape of the distributions vary significantly due to difference in the size of the sentences chosen for different languages. The sentences are short for English and Bengali, whereas long sentences are common in Indian languages, especially Tamil and Malayalam that are agglutinative. The QQ plot of the total number of syllables in a sentence with Gamma distribution for all languages is given in Figure 8. It is seen that all the data points lie more or less on a straight line thus indicating that they follow a Gamma distribution.

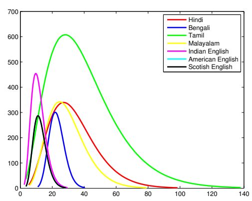

Figure 7: Gamma distribution of the total number of syllables

It is interesting to note that irrespective of languages, either stress-timed or syllable-timed, the total number of syllables in sentences and the number of syllables in any prosodic phrase follows a Gamma distribution.

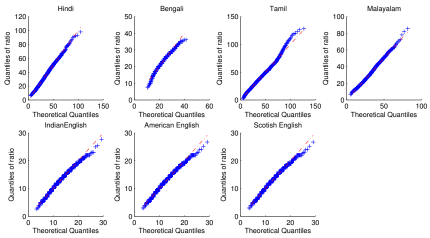

Figure 8: QQ plot of the total number of syllables in a sentence with Gamma distribution

3.2. Analysis of the ratio of the number of syllables

3.2.1. Ratio of number of syllables between all adjacent phrases

The histogram distribution of the ratio of number of syllables between each pair of adjacent prosodic phrases is analysed, and the tests of fit are performed with QQ plots. It is found that the data almost follows Gamma and Weibull distribution, but fits Gamma distribution better for all languages.The Gamma distribution is given in Figure 9(a) and the corresponding QQ plot is given in Figure 10. It can be seen from Figure 10 that the data points lies more or less on a straight line except for the tail of the distribution.

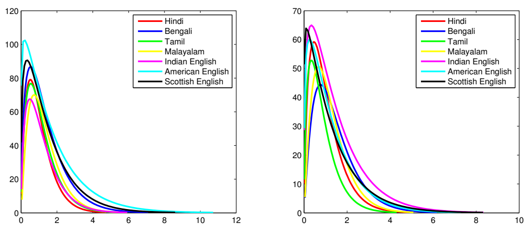

Figure 9: Gamma distribution of the ratio of the number of syllables between adjacent phrases (a) All adjacent phrases (b) 1st and 2nd phrase

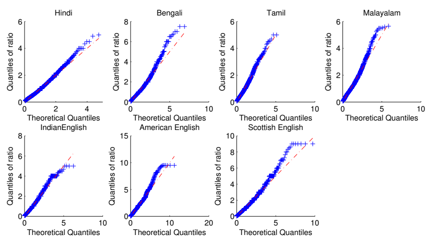

Figure 10: QQ plot of the ratio of the number of syllables between all adjacent phrases with Gamma distribution

3.2.2. Ratio of number of syllables for specific adjacent phrases

Ratio of number of syllables between specific pairs of adjacent phrases is analysed for the entire set of sentences and for different categories of sentences. The result observed is similar to that of the ratio of all pairs of adjacent phrases. The Gamma distribution for the ratio between the first and second phrase for two phrase sentences are shown in the second plot in Figure 9(b), taking the same number of sentences for all langauges.

It is very interesting to note from Figure 9(a) and 9(b) that the Gamma distribution of the ratio of number of syllables between different pairs of adjacent phrases is similar across all languages in terms of its shape and scale parameters, the parameters being, on an average, around 1.65 and 0.81 for Figure 9(a) and 1.67 and 0.87 for Figure 9(b). This is true irrespective of the length of the sentences or the distribution of number of syllables in different phrases.

4. Conclusion and Discussion

The objective of this paper is to understand syllable rhythm if present in a language. Experiments were performed to analyse the syllable rhythm in terms of the number of syllables in individual prosodic phrases and the ratio of number of syllables between adjacent prosodic phrases of different stress-timed and syllable-timed languages. Interestingly, it is observed that the syllabic rhythm at the phrase level is more or less language independent. It is observed that the syllable rhythm analyses using different characteristics of syllables uniformly follows either a Gamma or Weibull distribution for declarative sentences.

This suggests that syllabic rhythm is perhaps related to the vocal apparatus rather than the language. In [18] it is stated that during speech inhalation pause occurs at phrase or sentence boundaries. It is reported in [19] that there exists a positive correlation between inspiration duration and utterance length in terms of number of syllables for sentences uttered in a single breath. The syllabic rhythm at the prosodic phrase level could be related to the breathing pattern in humans[1].

The prosodic phrasings obtained were applied to TTS systems for synthesis of held-out sentences for two languages, Hindi and Tamil. Pair comparison (PC) listening tests were conducted on 12 listeners on a set of 16 natural and synthesized sentences. Listeners were asked to listen to the synthesized and original sentences in terms of naturalness with respect to the position of pauses, and give their preference. Table 3 shows the result of the PC test. In A-B test, the sentences in the system A is played first and in the B-A test, the sentences in the system B is played first. For both the languages, system A is the synthesized sentence and system B is the corresponding original sentence.

Table 3: *Results of pairwise comparison tests for Hindi and Tamil.*

Language	A-B (%)	B-A (%)	A-B + B-A (%)
Hindi	45.83	33.33	56.25
Tamil	65	60	52.9

It is observed that listeners did not perceive much difference between the original sentence and the synthesised sentence in terms of naturalness.

[1] We are at the time of this writing, not sure why this behaviour of rhythm is universal. We conjecture that it must be related to the production process (syllables being the fundamental unit of production). We are currently studying the physiology of speech production to see if such relationships indeed exist.

5. References

[1] D. Abercrombie, "Studies in Phonetics and Linguistics," *London: Oxford University Press*, 1965.

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

[3] K. Pike, "The Intonation of American English," *2nd edition. Ann Arbor: University of Michigan Press*, 1946.

[4] G. Kochanski, A. Loukina, E. Keane, C. Shih, and B. Rosner, "Long-range prosody prediction and rhythm," *Proceedings of Speech Prosody*, pp. 1-4, 2010.

[5] R. Dauer, "Stress-timing and syllable-timing reanalyzed," *Journal of Phonetics*, pp. 51-62, 1983.

[6] G. Esther, and E. Ling, "Durational Variability in Speech and the Rhythm Class Hypothesis," *Papers in laboratory phonology 7*, pp. 515-546, 2002.

[7] M. Khodabin, A. Ahmadabadi, "Some properties of generalized gamma distribution," *Mathematical Sciences 4.1*, pp. 9-28, 2010.

[8] Quantile-Quantile (q-q) plots, `http://onlinestatbook.com/2/advanced_graphs/q-q_plots.html`.

[9] P. G. Deepa, S. G. Veena, and A. S. Nair, "Modeling of Vowel Duration in Malayalam Speech using Probability Distribution," *Proceedings of IEEE, TENCON*, 2008.

[10] Indian Languages, `http://mhrd.gov.in/sites/upload_files/mhrd/files/upload_document/languagebr.pdf`.

[11] C. Fery, "Indian languages as intonational phrase languages" *I. Hasnain and S. Chaudhury [Eds] Festschrift to honour Ramakant Agnihotri*, Aakar Publisher, Delhi. 2010.

[12] Festvox: CMU-ARCTIC speech synthesis databases, `http://festvox.org/cmu_arctic/`.

[13] M. V. Vinodh, A. Bellur, D. M. Thakkare, A. Susan, H. A. Murthy, and others, "Using Polysyllabic units for Text to Speech Synthesis in Indian languages," *National Conference on Communication*, 2010.

[14] W. Fisher, "The tsylb2 Program: Algorithm Description. NIST. Part of the tsylb," 1996.

[15] A. W. Black, K. Lenzo, and V. Pagel "Issues in building general letter to sound rules," *International Speech Communication Association*, 1998.

[16] S. Aswin, "A hybrid approach to segmentation of speech using group delay processing and HMM based embedded reestimation," *INTERSPEECH*, 2014.

[17] H. Schmid, and M. Atterer "New statistical methods for phrase break prediction," *Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics*, 2004.

[18] G. Bailley, A. R. Capellan, and C. Vilain "Adaptation of respiratory patterns in collaborative learning" *14th Annual Conference of the International Speech Communication Association (INTERSPEECH)* 2013.

[19] D. H. Whalen, and J. M. Kinsella-Shaw, "Exploring the relationship of inspiration duration to utterance duration" *Phonetica* , pp. 138–152, 1997.