Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

The invalidity of rhythm class hypothesis

Chengxia Wang[1], *Yi Xu*[1], *Jinsong Zhang*[2]

[1]University College London, UK
[2]Beijing Language and Culture University, China
chengxia.wang.15@ucl.ac.uk, yi.xu@ucl.ac.uk, jinsong.zhang@blcu.edu.cn

Abstract

Languages are said to be stress-timed, syllable-timed or mora-timed. In a stress-timed language, inter-stress intervals are or tend to be constant, hence, isochronous, while in a syllable-timed or mora-timed language, successive syllables or morae are or tent to be equal in duration. Empirical research has failed to find evidence of isochrony in any language, yet the hypothesis is now sustained by perception accounts or phonetic metrics that do not measure isochrony. We have re-examined the rhythm class hypothesis by looking for evidence of at least a tendency toward isochrony, through a comparison of English, an alleged stress-timed language, and Mandarin, an alleged syllable-timed language. The results show that in English, segments are not compressible to allow equal syllable duration, and syllables are incompressible to enable equal inter-stress interval duration and phrase duration. In contrast, Mandarin shows a small tendency toward both equal syllable duration and equal phrase duration. These findings are exactly the opposite of what would be predicted by the rhythm class hypothesis. We therefore argue that the hypothesis is not just flawed, but simply untenable, and the so-called rhythm classes should no longer be held as a basic fact of human language.

Index Terms: rhythm class hypothesis, isochrony, compressibility, segment duration, syllable duration

1. Introduction

It has been widely accepted that languages of the world are either stress-timed, syllable-timed or mora-timed [1, 2, 3, 4, 5]. In a stress-timed language, inter-stress intervals are constant, hence, isochronous, whereas in a syllable-timed or mora-timed language, successive syllables or morae are equal in duration [1, 3, 4]. Experimental investigations, however, have been unable to find evidence of synchrony in either stressed-timed [6, 7, 8, 9, 10], syllable-timed [11, 12, 13] or mora-timed languages [14].

To achieve isochrony, the components of the isochronous units need to be flexible in duration, as recognized by [3, 4, 8]. For syllables to be isochronous, segmental duration needs to be flexible to compensate for changes in the number of segments in a syllable; and for stress groups to be isochronous, syllable duration needs to be flexible to accommodate the number of syllables in a stress group.

The frustration with the lack of evidence for isochrony seems to have been brushed away, however, by the proposal of the rhythm metrics since the late 1990s, starting from Ramus et al. [5]. These metrics are shown to be able to quantify the rhythm class of languages using consonantal and vocalic variability. The main measurements are %V, ΔV, ΔC [5], VarcoC, VarcoV [15, 16], and the pairwise variability indices nPVI and rPVI [17]. In recent years, researchers have used these rhythm metrics to study a wide range of languages and accents [18, 19, 20, 21, 22]. However, criticisms of the rhythm metrics quickly followed, including computational issues, their instability because of speech rate, speaking style, within-speaker variation, and measurement uncertainty, and their failure to clearly classify languages into alleged rhythm classes [18, 23, 24, 25, 26, 27].

The dissatisfaction with the rhythm metrics and the lack of evidence for isochrony has led to the suggestion that rhythm is a perceptual phenomenon rather than a fact of speech production [28, 29, 30]. Lehiste [31] shows that durational differences smaller than 30 ms are never reliably identified and concludes that "sentences that are not produced with absolutely isochronous intervals between stresses may still be perceived as if the interstress intervals were identical." But durational differences between inter-stress intervals can easily exceed 30 ms [30]. More critical for the rhythm class hypothesis, however, is whether listeners can consistently determine whether a language is syllable-timed, stress-timed or mora-timed. Not only is there no clear evidence that people have this ability, but also the classification by naïve listeners deviate from the rhythm class hypothesis more than trained phoneticians who are biased by the knowledge of the hypothesis [32] Miller. It was found that listeners could distinguish languages between and within rhythm classes [28, 33].

For the rhythm class hypothesis to remain tenable, we must go back to its core assumption, namely, isochrony of corresponding rhythmical units—stress, syllable or mora. Even if strict isochrony is not present, as is already known, there could exist a tendency toward isochrony. Such a tendency can be revealed by controlling all the other, non-rhythmic, factors that also affect timing and duration. These include lexical stress [34, 35], boundary marking, and intrinsic duration of segments [35].

Some research is already done in this direction, although not as a result of looking for evidence of isochrony. Multiple studies show that syllable duration in English increases quasi-linearly with syllable size, i.e., the number of constituent segments [9, 36]. Van Santen and Shih [37] showed that syllable duration is highly predictable from segmental duration in English, i.e., with every increment in the intrinsic duration of segments, syllable duration increases by almost the same amount. While this may be evidence of lack of syllable timing in English, it also makes it unlikely that syllable duration is flexible enough to achieve stress-timing, as discussed earlier. Interestingly, however, in the same study, syllable duration is not as highly correlated with vowel duration in Mandarin as in English, although the overall correlation is high between segmental and syllabic duration in Mandarin. Again, while this

10.21437/SpeechProsody.2022-71

may be a sign of segmental compression as evidence of syllable timing in Mandarin, it also opens up the possibility that the flexible syllable duration can be used to achieve some kind of equal duration of units larger than syllables. This interpretation is not contemplated in [37], however. Also, since only one male speaker was examined in each language in that study, the generalizability of the findings is not yet clear.

Findings of other studies indeed suggest that lack of flexibility in syllable duration may reduce the chance of isochrony of stress intervals. Inter-stress interval duration increases linearly with the number of constituent syllables in reiterant speech by English speakers [30]. A similar linear relationship between the number of intervening unstressed syllables and the inter-stress interval for real words in sentence context is also found by Lea [38] for English. These findings therefore show that in English, syllables are probably not compressed to maintain equal inter-stress intervals as the size of the inter-stress interval increases.

Nakatani et al. [30] find that lexical stress and position in word and phrase both have clear effects on syllable duration, but the effects are parallel to each other. For the positional effect, word-medial and word-initial syllables have largely the same duration, and the duration of word-final syllable is largely the same whether the word is monosyllabic or multi-syllabic. Xu and Wang [39], however, have found in Mandarin that phrase-medial syllables are shorter than phrase-initial syllables, and phrase-final syllables in multi-syllabic phrases are shorter than mono-syllabic words. Compared to English, Mandarin therefore has two additional means to shorten phrases as their sizes increase. This makes it likely that Mandarin has a tendency toward isochrony of phrases, contrary to the widely held belief that Mandarin is syllable-timed based on auditory impression and rhythm metrics analyses [17, 21, 40, 41].

With timing and isochrony back in focus, there seems to be evidence against the predictions of the rhythm class hypothesis in English and Mandarin. However, the findings have not led to a fundamental reconsideration of the rhythm class hypothesis, probably because the evidence is still rather scattered. To assess the generalizability of previous results, it is necessary to compare the duration patterns of example languages like English and Mandarin. The present study is a comparison of the timing patterns in two large non-experimental corpora, one in English and one in Mandarin, with the aim to both corroborate previous findings from controlled studies and answer further question critical for the rhythm class hypothesis by looking at the compressibility of segments and syllables in both languages.

2. Methods

As described in [42], the Boston University Radio News Corpus and [43] Annotated Corpus of Chinese Discourse [44] are used in current study.

2.1. Compressibility of segments

For syllables to show a tendency toward equal duration, their component segments must exhibit compressibility in one of two ways, or both. First, a segment would be compressed if its intrinsic duration is relatively long, so as to better match the intrinsically shorter ones. Our previous study shows that English segments are not compressed or stretched to make syllables equal in duration, while Mandarin tends to equalize syllable duration [42]. Second, all segments would be compressed as their numbers increase in a syllable.

For the relation of syllable duration and syllable size (number of component segments), a potential confound is that, in English, there is an uneven distribution of syllables of different sizes across boundaries of various strengths. Although the same trend is not seen in Mandarin, to avoid the bias it may bring, in the following analysis, we included only syllables before prosodic word boundary in the analysis. Also excluded from the analysis are syllables with the neutral tone in Mandarin.

As can be seen from figure 1, as the number of segments increases, syllable duration increases almost linearly in English, although the rate of increase is reduced slightly in the most complex syllables (those consisting of 5 segments). In Mandarin, in contrast, the rate is substantially reduced starting from 2-segment syllables.

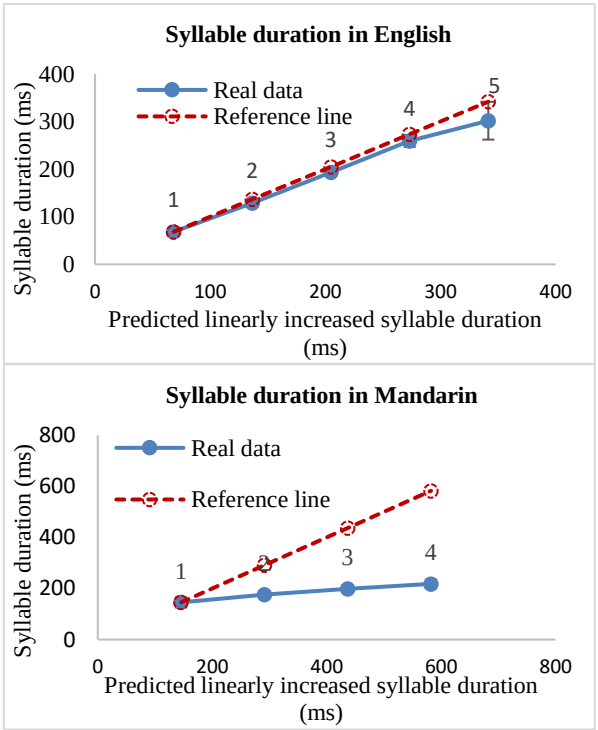

Figure 1: *Mean syllable duration in English and Mandarin as a function of syllable size (number of component segments) in stress, unstressed and all syllables. In each plot the dashed line is the reference line consisting of points each with the same increment as the duration of monosyllabic words (leftmost point).*

2.2. Compressibility of syllables

2.2.1. Inter-stress intervals in English

As shown in figure 2, inter-stress interval duration is highly related to interval size. The correlation between inter-stress interval duration and interval size is 0.981 ($p < 0.001$). Every unstressed syllable added increased inter-stress interval duration by 155 ms. This is consistent with previous findings [6, 10, 45, 46, 47].

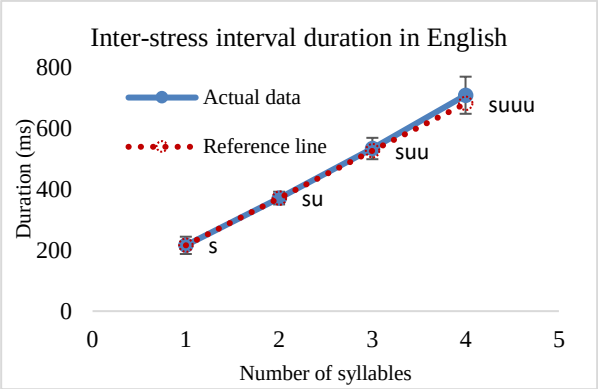

Figure 2. *Average inter-stress interval durations in seconds, s indicates a stressed syllable and u indicates an unstressed syllable. On the reference line, the leftmost point shows the duration of monosyllabic stressed syllable, while the rest of the points shows the same increment of the unstressed syllable in disyllabic intervals.*

2.2.2. Compressibility of syllable in prosodic phrases

To control for stress effect in English, the duration difference of each segment or consonant cluster is calculated and added to every segment and consonant cluster in unstressed syllable when computing phrase duration. By so doing, the shorter duration of unstressed syllables was not attributed to the reduction of phrase duration. Phrases with one or more neutral tone syllables in Mandarin were excluded from the analysis.

As can be seen in figure 3, phrase duration is strongly related to phrase size in both language: Pearson correlation coefficients are 0.984 ($p < 0.001$) in Mandarin and 0.987 ($p < 0.001$) in English.

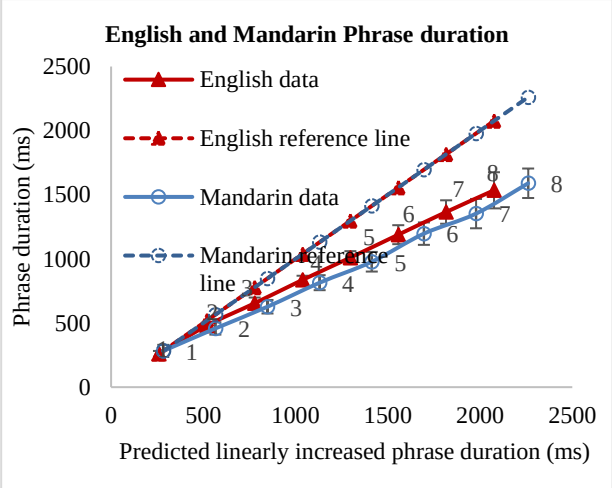

Figure 3. *Average duration of phrases in English and Mandarin as a function of phrase size in comparison with linearly predicted phrase duration.*

Figure 3 shows compression in both languages. But syllables in monosyllabic phrases are phrase final, which is subject to phrase-final lengthening. Using their mean duration as the baseline therefore provides an inflated reference slope, as phrase final lengthening does not apply to every syllable. To circumvent this problem, we then examined the compressibility of syllables in words as the number of syllables in a word increases. As shown in figure 4, word duration is strongly related to word size in both languages: Pearson correlation coefficients are 0.98 ($p < 0.001$) in Mandarin and 0.989 ($p < 0.001$) in English. But it can be also seen that syllables are compressed more in Mandarin than in English.

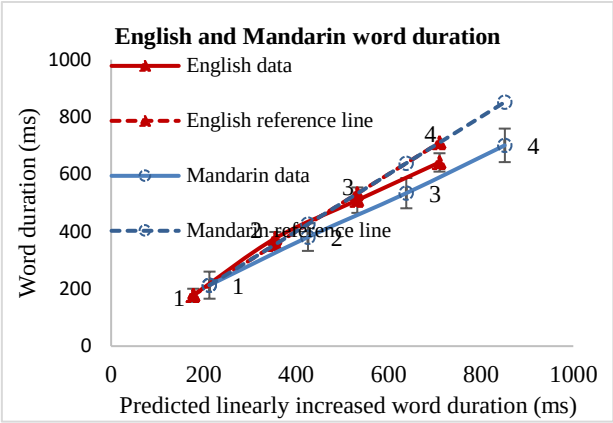

Figure 4. *Word duration in English and Mandarin as a function of predicted linearly increased word duration.*

To find out how syllables are compressed in words, we investigated syllable duration in terms of its position in word. Figure 5 shows how syllable duration depends on stress and position in word in English. Mixed Model ANOVAs were performed, with stress and position (word initial, word medial, word final) as fixed factors and syllable duration as dependent variable. The results showed a main effect of stress: $F(1, 5.123) = 159.150$, $p < 0.001$, partial $\eta 2 = .969$, and a main effect of position, $F(2, 11.621) = 10.005$, $p = 0.003$, partial $\eta 2 = .623$. Bonferroni post-hoc analyses showed that word final syllables are significantly longer than word initial and word median syllables. Although word initial syllables are slightly longer than medial syllables, they are not significantly different from each other. This is different from the findings of [30].

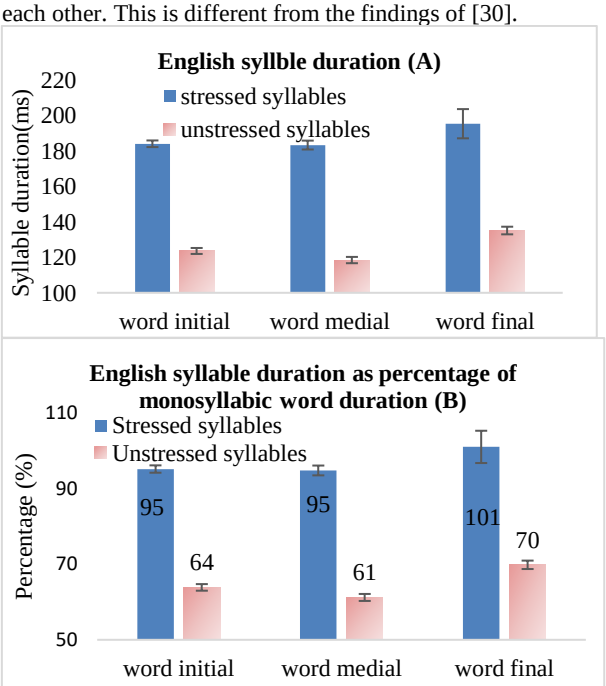

Figure 5. *English syllable duration as a function of position in word, in (A) milliseconds, and (B) percentage of monosyllabic word duration.*

Mixed Model ANOVAs were also performed on Mandarin data. The results showed a main effect of position, F (2, 21.101) = 160.133, p = 0.000, partial $\eta 2$ = .938. Bonferroni post-hoc analyses showed significant difference on each pairwise comparison between positions.

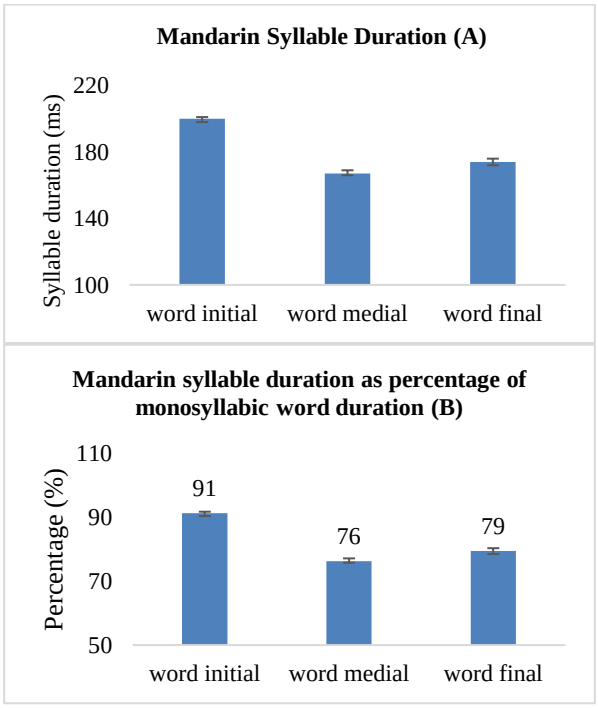

Figure 6. *Mandarin syllable duration as a function of position in word, in (A) milliseconds, and (B) percentage of monosyllabic word duration.*

Comparing Figures 5B and 6B, we can see that syllable duration is more compressible in Mandarin than in English. First, In English, word-final syllables are about equally long as monosyllabic words, whereas in Mandarin word-final syllables are much shorter than monosyllabic words. Second, in English, word-initial syllables are slightly but not significantly longer than word-medial syllables. In Mandarin, in contrast, word initial syllables are much longer than word medial syllables. The combined effects of word medial shortening and word final shortening, therefore, make Mandarin words much more compressible in duration than English words, for which both effects are absent.

What is also interesting is that word-final syllables in Mandarin are not only shorter than monosyllabic words, but also shorter than word initial syllables. Compared with mono syllabic words, word initial-syllables are 9% shorter, while word-final syllables are 21% shorter. This means that there is no word-final lengthening in Mandarin.

3. Discussion

The present study is a critical examination of the long-standing hypothesis that languages of the world are divided into stress-timed, syllable-timed and possibly mora-timed, each exhibiting a pattern of equal duration of the named units (Abercrombie, 1967; Pike, 1945). We have tried to go back to the most basic question about the hypothesis, namely, is there at least a tendency toward isochrony. This is done by comparing duration patterns of English, one of the few languages based on which the hypothesis was formulated in the first place, and Mandarin, a language that has been and confirmed as syllable-timed.

To the question of whether English shows a tendency toward isochrony at any level, the answer is no. Once various other duration-affecting factors are controlled, inter-stress intervals were found to linearly vary their duration with the number of constituent syllables (Figure 2). Furthermore, phrase duration and word duration also varied linearly with their size (Figure 3 and 4). This indicates that, in English, syllables are not compressible to show even a tendency toward equal inter-stress interval or equal phrase duration. This has removed the final trace of possibility that English is a stress-timed language based on isochrony. The reason for the incompressibility of English syllables has also become clear from our analysis. As shown in Figures 1, segments in English are also not compressible for the sake of equal syllable duration.

To the question of whether Mandarin shows a tendency toward isochrony, the answer is not only yes, but also that it happens at both the syllable level and the phrase level. First, syllables showed a tendency toward equal duration (figure 1). This is consistent with the prediction of syllable timing for Mandarin. However, it is also found that syllables are compressible for the sake of equal word duration or phrase duration in Mandarin (Figure 6). This is not predicted by the syllable-timing classification of Mandarin, although it is not directly against it either.

The absence of isochrony tendency in English, in contrast to Mandarin, begs an explanation. The direct cause, as already established, is likely the incompressibility of segments in the language. Previous research shows that it takes time to for articulators to move from current position to target position when articulating a segment [48]. Shortening syllables and hence segments beyond certain threshold would reduce intelligibility due to undershoot [49]. So lack of compression would help to maintain intelligibility. It is possible that functional load plays an important role here. Functional load was previously adopted to refer to the importance of certain phonetic contrast in languages [50]. Previous research shows that functional load of segments in English is larger than Mandarin [50], so it is more important to maintain minimal segmental duration in English than in Mandarin.

4. Conclusions

The results show an absence of any tendency toward isochrony of stress groups in English, but a small tendency toward both isochrony of syllables, isochrony of words and isochrony of phrases in Mandarin. Thus, the rhythm class hypothesis is argued to be not just weak, but untenable.

5. Acknowledgements

We thank Elizabeth Eden and Hao Liu for assistance with Python scripts, and Tristan Perryman for IT help and proof reading. This research was supported by China Scholarship Council to the first author.

6. References

[1] D. Abercrombie, “A phonetician’s view of verse structure,” *Linguistics*, vol. 2, no. 6, pp. 5-13, 1964.

[2] D. Abercrombie, “Syllable quantity and enclitics in English,” na, 1964.

[3] D. Abercrombie, Elements of General Phonetics, l967.

[4] K. L. Pike, The intonation of American English, 1945.

[5] F. Ramus, M. Nespor and J. Mehler, “Correlates of linguistic rhythm in the speech signal,” *Cognition*, vol. 73, no. 3, pp. 265-292, 1999.

[6] Y. Shen and G. G. Peterson, “Isochronism in English,” Department of Anthropology and Linguistics, *University of Buffalo*, 1962.

[7] E. T. Uldall, “Isochronous stresses in RP,” *Form and substance*, pp. 205-201, 1971.

[8] P. Roach, “On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages,” *Linguistic controversies*, 1982.

[9] J. D. O’Connor, “The duration of the foot in relation to the number of component sound-segments,” *Progress Report*, vol. 3, pp. 1-6, 1968.

[10] W. A. Lea, "Prosodic aids to speech recognition: IV," *A general strategy for prosodically-guided speech understanding. Univac Report PX10791, St. Paul, Minnesota: Sperry Univac*, 1974.

[11] B. J. Wenk and F. Wioland, “Is French really syllable-timed?” *Journal of phonetics*, 1982.

[12] A. M. Borzone de Manrique and A. Signorini, “Segmental duration and rhythm in Spanish.” *Journal of Phonetics*, **11**(2): 117-128, 1983.

[13] G. E. Pointon, “Is Spanish Really Syllable-Timed?” *Journal of phonetics*, *8*(3), 293-304, 1980.

[14] N. Warner and T. Arai, “Japanese Mora-Timing: A Review.” *Phonetica*, *58*(1-2), 1-25, 2001.

[15] V. Dellwo, “Rhythm and speech rate: A variation coefficient for ΔC,” *Language and language-processing*, pp. 231-241, 2006.

[16] V. Dellwo and P. Wagner, “Relationships between rhythm and speech rate,” *15th International Congress of the Phonetic Sciences*, Barcelona, pp. 471-474, 2003.

[17] E. Grabe and E. L. Low, “Durational variability in speech and the Rhythm Class Hypothesis,” *In Laboratory phonology 7* (pp. 515-546), 2008.

[18] A. Arvaniti, “The usefulness of metrics in the quantification of speech rhythm,” *Journal of Phonetics*, vol. 40, no. 3, pp. 351-373, 2012.

[19] J. Dankovičová and V. Dellwo, “*Czech speech rhythm and the rhythm class hypothesis.” Paper presented at the Proceedings of the 16th International Congress of Phonetic Sciences* (ICPhS), 2007.

[20] P. Mok, “On the syllable-timing of Cantonese and Beijing Mandarin,” *Chinese Journal of Phonetics*, vol. 2, pp. 148-154, 2009.

[21] F. Nolan and E. L. Asu, “The pairwise variability index and coexisting rhythms in language,” *Phonetica*, vol. 66, no. (1-2), pp. 64-77, 2009.

[22] E. O’Rourke, “Speech rhythm variation in dialects of Spanish: applying the pairwise variability index and variation coefficients to Peruvian Spanish,” *Speech Prosody*, 2008.

[23] P. M. Bertinetto and C. Bertini, “On modeling the rhythm of natural languages,” *In Proceedings of the Fourth International Conference on Speech Prosody*, 2008.

[24] D. Deterding, “The measurement of rhythm: A comparison of Singapore and British English,” *Journal of Phonetics*, vol. 29, no. 2, pp. 217-230, 2001.

[25] D. Gibbon, “Computational modelling of rhythm as alternation, iteration and hierarchy,” *In Proceedings of ICPhS*, vol. 15, 2003.

[26] R. A. Knight, R. A. “Assessing the temporal reliability of rhythm metrics,” *Journal of the International Phonetic Association*, vol. 41, no. 3, pp. 271-281. 2011.

[27] F. Nolan and H.-S. Jeon, “Speech rhythm: a metaphor?” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *369*(1658), 20130396, 2014

[28] A. Arvaniti, “Rhythm, timing and the timing of rhythm.” *Phonetica*, 66(1-2), 46-63, 2009.

[29] K. J. Kohler, “Rhythm in Speech and Language.” *Phonetica*, *66*(1-2), 29-45, 2009.

[30] L. H. Nakatani, K. D. O’Connor and C. H. Aston, “Prosodic aspects of American English speech rhythm,” *Phonetica*, vol. 38, no. (1-3), pp. 84-105, 1981.

[31] I. Lehiste, “Isochrony reconsidered,” *Journal of phonetics*, vol. 5, no. 3, pp. 253-263, 1977.

[32] M. Miller, “On the perception of rhythm.” *Journal of Phonetics*, 12(1), 75-83, 1984.

[33] L. White, S. L. Mattys and L. Wiget, “Language categorization by adults is based on sensitivity to durational cues, not rhythm class.” *Journal of Memory and Language*, 66(4), 665-679. 2012.

[34] D. B. Fry, “Experiments in the perception of stress.” *Language and speech*, 1(2), 126-152, 1958.

[35] D. H. Klatt, “Linguistic uses of segmental duration in English: Acoustic and perceptual evidence.” *Journal of the Acoustical Society of America* **59**: 1208-1221, 1976.

[36] T. H. Crystal and A. S. House, “Articulation rate and the duration of syllables and stress groups in connected speech.” *The Journal of the Acoustical Society of America*, 88(1), 101-112, 1990.

[37] J. P. Van Santen, and C. Shih, “Suprasegmental and segmental timing models in Mandarin Chinese and American English,” *The Journal of the Acoustical Society of America*, vol. 107, no. 3, pp. 1012-1026, 2000.

[38] W. A. Lea, “Isochrony and disjuncture as aids to syntactic and phonological analysis,” *The Journal of the Acoustical Society of America*, vol. 57, no. S1, pp. S33-S33, 1975.

[39] Y. Xu and M. Wang, “Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin,” *Journal of Phonetics*, vol. 37, pp. 502-520, 2009.

[40] H. Lin, and Q. Wang, “Mandarin rhythm: An acoustic study,” *Journal of Chinese Language and Computing*, vol. 17, no. 3, pp. 127–140, 2007.

[41] P. Mok, and V. Dellwo, “Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English,” *In Proceedings of Speech Prosody*, vol. 4, pp. 423-426, 2008.

[42] C. Wang, J. Zhang and Y. Xu, “Compressibility of segment duration in English and Chinese.” *In Proceedings of Speech Prosody 2018* (Vol. 9, pp. 651-655). International Speech Communication Association (ISCA), 2018.

[43] M. Ostendorf, P. J. Price, and S. Shattuck-Hufnagel, “The Boston University radio news corpus,” *Linguistic Data Consortium*, pp. 1-19, 1995.

[44] A. Li, X. Chen, G. Sun, W. Hua, Z. Yin, Y. Zu, F. Zheng, and Z. Song, “The phonetic labeling on read and spontaneous discourse corpora,” *ICSLP*, 2000.

[45] A. Classe, “The rhythm of English prose,” *B. Blackwell*, 1939.

[46] D. L. M. Bolinger, “Forms of English: Accent, morpheme, order,” *Harvard University Press*, 1965.

[47] J. D. O'Connor, “The perception of time intervals.” Progress Report 2, *London: Phonetics Laboratory*, University College.1965.

[48] P. Birkholz, B. J. Kroger and C. Neuschaefer-Rube, “Model-based reproduction of articulatory trajectories for consonant–vowel sequences.” *IEEE Transactions on Audio, Speech, and Language Processing*, 19(5), 1422-1433.2010.

[49] C. Cheng and Y. Xu, “Articulatory limit and extreme segmental reduction in Taiwan Mandarin.” *The Journal of the Acoustical Society of America*, 134(6), 4481-4495. 2013.

[50] D. Surendran and G. A. Levow, “The functional load of tone in Mandarin is as high as that of vowels.” *Speech Prosody*, 2004.