Speech is “heterometric”: The changing rhythms of speech

Steven Brown[1], Kyle Weishaar

[1] **Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario**

[2] **Department of Linguistics and Languages, McMaster University, Hamilton, Ontario**

stebro@mcmaster.ca, weishak@mcmaster.ca

Abstract

Work on speech rhythm has been notoriously oblivious to describing actual rhythms in speech. We present here a model of speech rhythm inspired by musical conceptions of meter. We posit that changes in meter are central to speech rhythm, and thus that speech is “heterometric” rather than isochronous. In addition, we see two devices for obviating the need for meter changes within a sentence, both of them involving subdividing component beats: 1) subdivisions according to simple integer ratios, resulting in duplets and triplets; and 2) subdivisions according to complex ratios, resulting in polyrhythms.

Index Terms: Speech, rhythm, meter, music, heterometer, polyrhythm, prominence, stress

1. Background

Work on speech rhythm has been driven more by a desire to classify languages into categories than by the need to elucidate the actual rhythms of spoken sentences. Contemporary approaches, for example, focus on the variability of syllable durations in sentences [1] or the proportion of sentences that is occupied by vowels [2]. But these features don’t specify actual rhythms. In fact, different metrics can sometimes classify the same language into opposing rhythm types ([3] & [4]). Even within a language, speech style can vary greatly due to dialect, sociolect, and speech rate [5]. Thus, language taxonomies only consider prototypical examples of language, and do not account for the actual rhythms of languages.

Classic work on “stress-timed” vs. “syllable-timed” languages ([6] & [7]) comes closer to describing the temporal features of a language. However, these models of isochrony are too simple, and make the assumption that the rhythm of a language is constant for all utterances ([8] & [9]) and all speakers. In addition, the idea that a language falls into a discrete category (e.g., syllable-timed, stress-timed) seems to be more of a perceptual phenomenon than a result of generative mechanisms, and much research has shown that perceptual regularity does not necessarily rely on stimulus regularity ([10],[11],[12] & [13]). Furthermore, the classic notions of stress-timing and syllable-timing ([14] & [15]) were based on subjective impressions of language rhythms, and empirical studies that measured the durations of inter-onset-intervals failed to find the type of isochrony proposed in these models.

Other important approaches to speech rhythm include metrical phonology and rhythm metrics, which describe rhythm in terms of the weight distribution of syllables throughout an utterance, as a function of phonological environment. These are powerful systems that explain stress shifts and prominence deletion quite well with language-specific rhythm rules [14]. But while this type of model accounts for the distribution of accentual prominence over an utterance, it says nothing about the *relative durations* of syllables in a sentence, a key consideration for any notion of rhythm.

2. Model

We offer a new, music-inspired approach to the analysis of speech rhythm, and propose that speech is “heterometric”. By this we mean that there are local pockets of metricality in sentences but that meters can change throughout the course of a sentence, for example from a triple-based meter to a duple-based meter. The concept of a heterometer, like that of a meter more generally, relies on notions of grouping. Grouping effects in music perception [13] and speech perception [15] are well-described phenomena. They allow the perceptual system to process organized chunks of information, as opposed to a linear piece-by-piece strategy.

The early models of speech isochrony implicitly assumed (usually without empirical evidence) that a single meter permeated all the sentences of a language, but we show here that speech is much more akin to music, and that local changes in rhythm occur, as in the case of heterometric musics in traditional cultures. While our approach in the lab has been to create musical models of sentences and test their rhythmic predictions using acoustic measurements of intervals between prominent syllables, the present discussion is of a more theoretical nature, employing musical transcriptions of sentences.

2.1 Prominence groups

We propose that the basic unit of speech rhythm is the “prominence group”, analogous to a bar or measure in music. The defining feature of a prominence group is that it always begins with a strong beat (i.e., a stressed syllable in the case of English), just as a musical measure always begins with a strong beat. This is formally analogous to the rhythmic units proposed in isochrony models of speech rhythm (although our groupings needn’t be isochronous; see below).

Let us consider a passage with obvious stress timing and meter, namely the verse "Hickory Dickory Dock" (Figure 1). Its rhythm is a simple triple meter, which implies an alternation between one strong beat and two weak beats having equal inter-onset intervals.

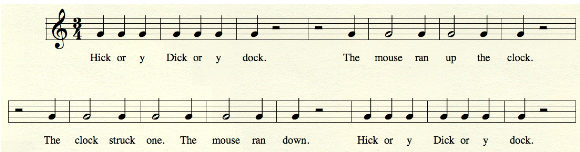

Figure 1. A nursery rhyme in triple meter.

Just as in any description of musical rhythm, each syllable here is assigned a duration value. Importantly, these are *relative* duration values, and an understanding of absolute tempo would require a specification of the duration of a note-value at some level. Syllables differ in their relative duration values, with some being twice the duration of others in the example given here (e.g., "mouse" vs. "ran"). Several factors contribute to variability in duration for syllables. One is phonemic content. Phonemes have inherent durational values that vary depending on their articulation and local environment [16]. The phonemic composition of a syllable contributes to its durational value. For example, the initial voiceless plosive /k/ will occupy less temporal space than the sibilant /s/. Likewise, consonant clusters can make syllables longer than simpler syllables (e.g., CCCVCCC vs. CV). Languages that are classified as stress-timed tend to have more-complex syllable structures than those classified as syllable-timed, and thus have more variability of syllable types and durations ([7] & [8]). Complexity in syllable structure creates a need for speakers to use processes such as reduction, deletion, and lengthening in order to maintain temporal regularity [8].

Given the presence of syllabic stress in English, stressed syllables are necessarily going to be those that align with the beginnings of prominence groups, e.g., the "hick" of "hickory". Note that simple syntactic constituents like noun phrases ("the mouse", "the clock") are often broken apart in these prominence groups, since prominence groups are rhythmic units rather than linguistic units. At the same time, the contrast between content and function words is important, in that content words are the ones that are most likely to head prominence groups, whereas function words are the ones that are most likely to undergo durational reductions (for example, clitics). Hence, the arrangement of strong and weak beats in a prominence group is influenced by morphological and syntactic constraints, such as compounding and placement of function words, but does not need to necessarily correspond with morphological or syntactic units.

2.2 Heterometers: Changes of meter within a sentence

A natural sentence spoken by an individual will not have the rhythmic simplicity of a passage of composed verse. A significant departure of our model from classic models of isochrony is that it posits the occurrence of meter changes within sentences, e.g., from a triple meter to a duple meter. Hence, we propose that speech is heterometric, and that meter-change is a central feature of speech rhythm, especially in longer or more-complex sentences. Consider the sentence "Nathaniel writes novels and lives in a green house built by a farmer":

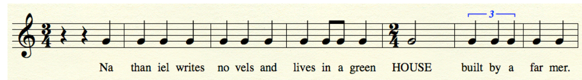

Figure 2. Demonstration of a meter change within a sentence, exemplifying a heterometer.

The initial part of the sentence – "Nathaniel writes novels and lives in a green" sets up a basic triple meter. The second phrase – "house built by a farmer" – if it were it to stand alone would be best modeled as a duple meter. Its first prominence group is a result of syntactic stress, with the noun "house" carrying the focus of the phrase. Putting the two phrases together, we see the change in meter from triple to duple. It's quite easy to construct sentences that have the reverse pattern. In general, we have found that 2- and 3-beat meters serve as the basic building blocks of speech rhythm, just as they do for musical rhythm.

When meter-changes occur (and sometimes even when they don't), the tempo can change as well. In other words, the durational value of the basic beat can become shorter or longer. Hence, another important feature of speech rhythm is not only changes in the metric groupings across a sentence but also changes in the duration-value of the beats within that meter, in other words *tempo change*.

It is too early to elaborate all the factors that contribute to meter, meter change, and tempo change in sentences, but we believe that the study of speech rhythm should be dedicated to a search for these principles. While contemporary models are driven by taxonomic concerns for grouping languages into categories, it seems to us that once these rhythmic principles are understood, grouping should be a relatively straightforward exercise. However, the converse will probably not be the case.

2.3 Simple subdivisions of beats: Duplets and triplets

A reasonable optimality rule for speech rhythm would be to minimize meter changes within a sentence. To this end, we can imagine two principal meter-preserving mechanisms in speech. Both of them involve creating subdivisions of the basic beat and thus generating a metrical hierarchy for the phrase: 1) subdividing beats according to simple integer ratios to generate duplets and triplets, and 2) subdividing beats in a complex fashion to generate polyrhythms (discussed in section 2.4 below).

One weakness of the classic isochrony models is that they don't talk about units of duration at the syllabic level ([6] & [7]), whereas contemporary models talk about durational variability of syllables and about proportions of an utterance occupied by vowels, but without specifying actual rhythms ([1] & [2]). Therefore, we take advantage of the analogy to music to specify not only units of duration within simple metric groupings but also *subdivisions* of these beats, resulting in durational variability of syllables throughout a sentence. (For a similar argument for poetry, see [17]). In music's metrical hierarchy, these subdivisions generally take the form of small integer ratios, such as duplets (each one having one half the duration of the basic beat) and triplets (each one having one third the duration of the basic beat), and this can be shown to be the case in speech as well. This is seen above in the phrase "in a" and the word "farmer" (duplets) as well as in the phrase "built by a" (a triplet). An alternative version would have "built by a" with the same rhythmic pattern as "lives in a".

Such duplets and triplets reflect the fact that syllable durations are *compressed* in speech. For languages like English, there are well-characterized phenomena like vowel reduction that lead to corresponding reductions in syllable duration. Likewise, certain function words, such as clitics, are monosyllabic words that tend to get uttered in a very reduced manner. Hence, both syllable stress and syntactic role become factors in defining compressions in syllable duration [18].

2.4 Complex subdivisions of beats: Polyrhythms

A second meter-preserving rhythmic device of subdivision is polyrhythm. In music, the concept of polyrhythm implies a conflict between incompatible rhythms. For example, if two people were to simultaneously tap a 3-beat and 2-beat rhythm, respectively, against the same drumbeat, this would create a 3-against-2 polyrhythm, since 3 and 2 are not divisible by a common integer factor (except 1). Alternations between 2-beat and 3-beat figures in music are referred to as "hemiolas", and this seems to be a common feature of polyrhythm in speech. Polyrhythm is another manifestation of the phenomenon of subdivision, but one in which the beats are not mutually divisible as simple integer ratios.

Here we present the same sentence as above but convert it into its compound-noun form, i.e., change the adjectival "green house" into the compound "greenhouse": "Nathaniel writes novels and lives in a greenhouse built by a farmer" (Figure 3). Notice that this change in nuclear stress results in a dramatic change in sentence rhythm, since the prominence groups have to be readjusted for morphological stress:

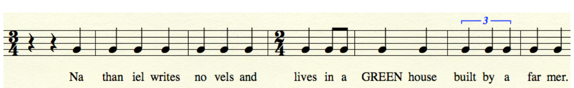

Figure 3. A change in nuclear accent leads to a change in sentence rhythm. Compare with Figure 2. One result is a polyrhythm for "built by a".

In this sentence, "built by a" can be represented as a polyrhythm (i.e., 3 beats against the 2 beats of the duple rhythm of the phrase). Different renditions of this sentence by different speakers could also model it with the same durational pattern as "lives in a". However, the next sentence would not have such flexibility, and could only be modeled in a purely polyrhythmic fashion (Figure 4): "Pamela purchased beautiful flowers Saturday morning all through the year".

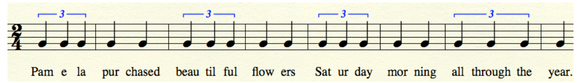

Figure 4. A completely polyrhythmic sentence. Also, an example of a hemiola in speech.

2.5 Other considerations

Work on speech cycling, synchrony, and perceptual centers by Cummins and by Port ([19],[20] & [21]) shows that certain rhythmic patterns of speech are permissible, while other are less so. Neural oscillations attract the onset of syllables, whether these be stressed or non-stressed syllables [21]. There is evidence for attractors operating at harmonic fractions of beats, especially halves and thirds [21]. This supports our notion of duplets and triplets as being basic subdivisions within speech's metrical hierarchy. Port proposes that these oscillators attract perceptual attention and influence the production of sentences in the motor system of the brain. This mechanism is what underlies meter. Both Port and Cummins claim that this process has a coupling effect such that perception of beats and production of beats align to a pulse. Such pulses can be either external, such as metronome clicks, or internal, via neural oscillators ([20] & [21]). If this proposal is correct, then both the perception of regularity in speech and the production of metrically-recurring beats can be accounted for in our model of speech rhythm. For other important oscillator-based approaches to speech rhythm, see [22], [23], and [24].

3. Cross-linguistic considerations

Preliminary analysis of Cantonese rhythm bolsters the findings presented here for English [25]. A meter can be established based on isochrony of syllables. In Cantonese, sentences seem to end obligatorily on strong beats, hence leading to a "downbeat rule" for sentence generation. Subdivisions of beats ("compressions") occur that reduce durational values, doing so according to small integer ratios, such as duplets and triplets. These seem to be guided quite strongly by syntactic considerations. Meter changes and polyrhythms are likely to occur in Cantonese, since phrases are constructed in duple and triple meters. Sentence construction will then lend itself to meter changes and perhaps polyrhythms, although less frequently than in English.

What will be the determinants of these rhythmic mechanisms cross-linguistically? At least two inter-dependent factors seem to be strong candidates: polysyllabicity of words and the presence of syllabic stress within words. Languages like English that have polysyllabic words with word stress probably lend themselves to having meter changes in sentences. Languages that are more monosyllabic will probably have more constant meters. But even a language like Cantonese that has a simpler syllable structure than English, and is thus less prone to meter change, still shows subdivisions of beats in a pervasive manner. Hence, subdivision of beats might be a more general rhythmic mechanism than heterometers.

In our opinion, the classic dichotomy between stress-timed and syllable-timed languages is in serious need of an overhaul. Speech rhythm seems to be inherently based on stress timing, even for languages like Cantonese that lack word-level stress. What seem to vary across languages are the kinds of features we have talked about: the durational variability of constituents that sit between stress points (i.e., subdivisions); the presence of meter changes; and the presence of tempo changes. We suspect that there is no language that is based on constant strings of isochronous syllables. Instead, one should find, at one end of the spectrum, rhythmically simpler languages that have few subdivisions of beats, relatively constant meters, and relatively constant tempos. At the other end should be rhythmically complex languages that have greater numbers of subdivisions of beats, more frequent meter changes, and more frequent tempo changes. From our limited experience with this analysis, Cantonese and English might represent prototypes, respectively, of these two varieties of speech rhythms. This jibes perfectly with the well-established notion that languages differ in the durational variability of their syllables [1].

4. Summary of model

Our heterometric model posits a few fundamental rhythmic mechanisms that should be applicable across languages. We see a basic similarity of speech rhythm to the structure of music through an organization of sentences into prominence groups headed by strong beats. Next, we posit that meter-change is central to speech rhythm, and thus that speech is heterometric rather than isochronous. Tempo changes can also occur during the course of a sentence, altering the duration values of beats. In addition, we see two meter-preserving rhythmic mechanisms involving subdivisions of beats: 1) subdividing according to simple integer ratios to generate duplets and triplets; and 2) subdividing according to complex ratios to generate polyrhythms. While the relative importance of these diverse mechanisms varies across languages, it is likely that all of these mechanisms are present in some form in all languages.

5. References

[1] Grabe, E. & Low, L. 2003. Durational variability in speech and the rhythm class hypothesis. In N. Warner & C. Gussenhoven (eds.), *Papers in Laboratory Phonology* 7. 515-546. Berlin: Mouton de Gruyter.

[2] Ramus, F., Nespor, M. & Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73. 265-292.

[3] Bertinetto, P. M. 1989. Reflections on the dichotomy 'stress' vs. 'syllable-timing'. *Revue de Phonétique Appliquée* 91/93. 99-130.

[4] Arvaniti, A. In press. Rhythm, timing, and the timing of rhythm. *Phonetica*.

[5] Russo, M. & Barry, W. 2008. Isochrony reconsidered: Objectifying relations between rhythm measures and speech tempo. In *Proceedings of Speech Prosody,* 419-422. Campinas, Brazil.

[6] Pike, K. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.

[7] Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

[8] Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11. 51-62.

[9] Faure, G, Hirst, D. J. & Chafcouloff, M. 1980. Rhythm in English: Isochronism, pitch and perceived stress. In L. R. Waugh and C. H. van Schoolneveld (eds.), *The Melody of Language,* 71-79. Baltimore: University Press Park.

[10] Uldall, E. T. 1971. Isochronous stresses in RP. In E. Hammerich (ed.), *Form and Substance*, 205-201. Copenhagen: Akademisk Forlag.

[11] Lehiste, I. 1977. Isochrony reconsidered. *Journal of Phonetics* 5. 253-263.

[12] Donovan, A. & Darwin, C. J. 1979. The perceived rhythm of speech. In *Proceedings of the 9th International Congress of Phonetic Science*, vol. 2, 268-274. Copenhagen: Institute of Phonetics, University of Copenhagen.

[13] Drake, C. & Palmer, C. 1993. Accent structure in music performance. *Music Perception* 10(3). 343-378.

[14] Nespor, M. & Vogel, I. 1986. *Prosodic phonology.* Dordecht: Foris Publications.

[15] Boucher, V. J. 2006. On the function of stress rhythms in speech: Evidence of a link with grouping effects on serial memory. *Language and Speech* 49(4). 495-519.

[16] Peterson, G. E. & Lehiste, I. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32(6). 693-703.

[17] Lerdahl, F. 2001. The sounds of poetry viewed as music. *Annals of the New York Academy of Sciences* 930. 337-354.

[18] Chomsky, N. & Halle, M. 1968. *The sound pattern of English*. New York: Harper and Row.

[19] Cummins, F. In press. Rhythm as entrainment: The case for synchronous speech. *Journal of Phonetics*.

[20] Cummins, F. & Port, R. 1998. Rhythmic constraints of stress timing in English. *Journal of Phonetics* 26. 145-171.

[21] Port, R. 2003. Meter and speech. *Journal of Phonetics* 31. 599-611.

[22] O'Dell, M. L. & Nieminen, T. 1999. Coupled oscillator model of speech rhythm. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* 2. 1075–1078. University of California, Berkeley.

[23] Saltzman, E., Nam, H., Krivokapic, J. & Goldstein, L. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa, S. Madureira & C. Reis (eds.), *Proceedings of the 4th International Conference on Speech Prosody* (Speech Prosody 2008), Campinas, Brazil.

[24] Tilsen, S. 2009. Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33. 839-879.

[25] Chow, I., Brown, S., Poon, M. & Weishaar, K. A model for rhythmicity in spoken Cantonese. (Paper submitted for this conference).