Speech Rhythm

Dafydd Gibbon (2021-01-28)

Rhythm is easy to recognise. One can feel it, hear it and see it, in music, in dance, in walking and running, in speech, in singing, in heartbeats, in the ticking of a clock. But it is harder to define rhythm than to recognise it. Is a rhythm a beat, or a wave? Is it completely regular or can it be syncopated? How many kinds of rhythm are there? How are rhythms in music, poetry and speech related? Are rhythms the same for humans and animals? Is the regularity of rhythm based on points in time as beats or intervals in time as waves? What are the frequencies of rhythms? If a regular beat or wave is very fast, for example more than ten per second, or very slow, for example fewer than one every five seconds, is it still a rhythm? Answers to these questions are sought in scientific models and methods of analysis. A basic scientific model of rhythms, including a speech rhythm, can be thought of as an oscillating event series, and interacting simultaneous speech rhythms can be modelled by coupling the oscillators which generate these rhythms (Cummins and Port 1998, Barbosa 2002).

In spoken language, speech, there are many factors which can contribute to the complexity of these coupled oscillations, and languages have different properties which make the observable complex rhythms sometimes quite different: differences in *word structure* (languages with and without prefixes and suffixes, like English, in contrast to Chinese); differences in *sentence structure* (languages with and without a pervasive distinction between grammatical and content words, again like English and Chinese, or which have verbs at the beginning, in second place or at the end of sentences, like Welsh, English and German, respectively); the patterns of contrast and emphasis associated with introducing new *information* into a discourse (*JOHN wore a RED shirt and FRED wore a BLUE one*). In poetry, these factors interact with conventions about *metre*, the clearly defined rhythmic patterns which characterise different kinds of poem. For example, grammar and metre may conflict, as in *I WANdered LONEly as a CLOUD* according to grammar, but *I WANdered LONEly AS a CLOUD* according to the iambic metre of a weak-strong syllable sequence. In *orthography*, rhythms are sometimes rendered with highlighting, as in the examples given here, or sometimes, in social media, with dots: *This.is.definitely.true!*

It is not easy to define rhythm, as already noted, but it is even harder to analyse the detailed rhythms of speech scientifically. There have been three main approaches to the scientific analysis of rhythm. The first, dating back to the early 20th century (Jones 1918), is a qualitative pedagogical approach, which identified differences between languages like English, with fairly regularly timing (*isochrony*) of sequences of *rhythm units* (stressed syllables and intervening unstressed syllables), and languages like French, with fairly regularly timed syllables as the rhythm units. These basic units of rhythm are organised into *rhythm groups* which in general correspond to the constraints of grammar and information structure (cf. Palmer 1924, Pike 1946, Jassem 1952, Abercrombie 1967). From the early 20th century until today this distinction between *stress timing* and *syllable timing*, which has also been said to characterise the difference between European and Brazilian Portuguese (cf. Barbosa et al. for discussion), has been standardly used in language teaching.

In the second half of the 20th century, the syllable-stress timing distinction came to be seen as controversial and formed the starting point for a second approach: quantitative studies of the durations of different kinds of rhythm unit (Jassem et al. 1984, Roach 1982, Ramus et al. 1999, Low

et al. 2000, Asu and Nolan 2006). Using descriptive statistics (average, variance and related measures), these studies were able to show that languages differ fairly systematically along a scale of regularity and irregularity of the duration patterns of syllables and other units in sentences, even though there is no isochrony in the strict sense. For example, using one of these measures (Low et al. 2000), the *nPVI (normalised Pairwise Variability Index*), it can be shown that English typically has a value around 60, indicating higher irregularity, while Chinese has a value around 35, indicating lower irregularity (zero would indicate total regularity, i.e. isochrony). In these statistical approaches, linguistically defined rhythm units such as consonantal or vocalic segments, or syllables, are assigned time-stamps to denote their start and finish points in the speech recording; from the time-stamps durations of the segments are calculated, and average differences in duration or variance of duration are calculated. In Figure 1, pronunciations of the two versions of a sentence, *Astrid SANG very WELL* (left) and *AStrid sang very WELL*, right, are shown, spoken one after the other, together with traces of the intensity and the fundamental frequency (F0, 'pitch') of the signal.

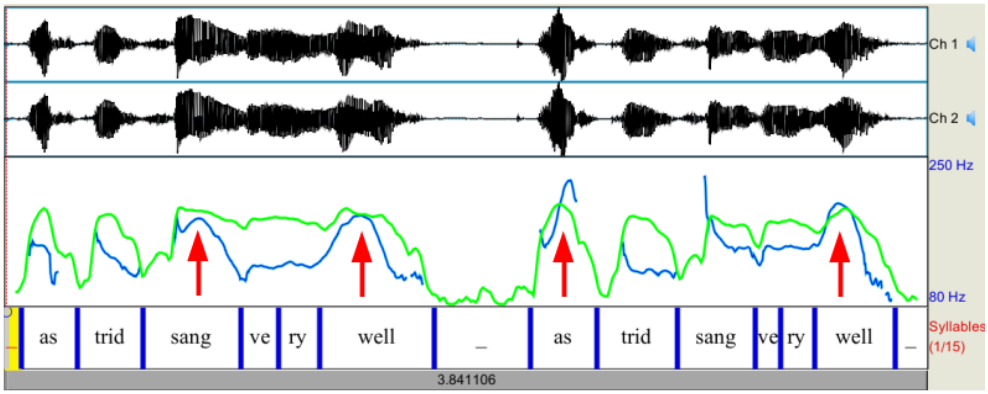

Figure 1: Annotated speech signal with intensity track (green) and fundamental frequency track (blue): (1) left, grammatical pattern; (2) right, information structure pattern. Positions of accented words are marked with arrows. Software: Praat (Boersma 2001).

Some properties of the two utterances are immediately obvious. In the main accented syllables *sang* and *well* in the first utterance, and *as* and *well* in the second, the high intensity (green curve) and high fundamental frequency (blue curve) are very clear. Duration relations are not so clear: in the first example the accented syllables *sang* and *well* are longer than the other syllables, but in the second example unaccented *strid* is longer than accented *as*.

The statistical studies have been useful in pointing out that syllable timing in some languages is more irregular than in others, but they have not addressed the core issues of rhythms as oscillations with frequencies. They also have a methodological problem: while descriptive statistics are suitable for describing static populations, they are not suitable for describing time series, of which rhythms are an example, which require signal processing and probability measures.

From the late 1990s a third approach emerged, which used signal processing methods and analysed speech rhythms as modulations of a basic carrier signal produced in the larynx, concentrating mainly on the amplitude modulations produced by the 'waves' or 'beats' of syllables and other rhythm units, which result from alternations of lower amplitude consonants and higher amplitude vowels (Todd et al. 1994, Galves et al. 2002, Tilsen and Johnson 2008). The signal processing approach starts with a simple fact: rhythms have frequencies. Using basic signal processing techniques, this approach looks for low frequency changes in the amplitude contour or *envelope* of the signal, which are below about 10 Hz, that is, about 10 beats or waves per second. This can be shown in a rhythmical counting sequence from *one* to *ten*.

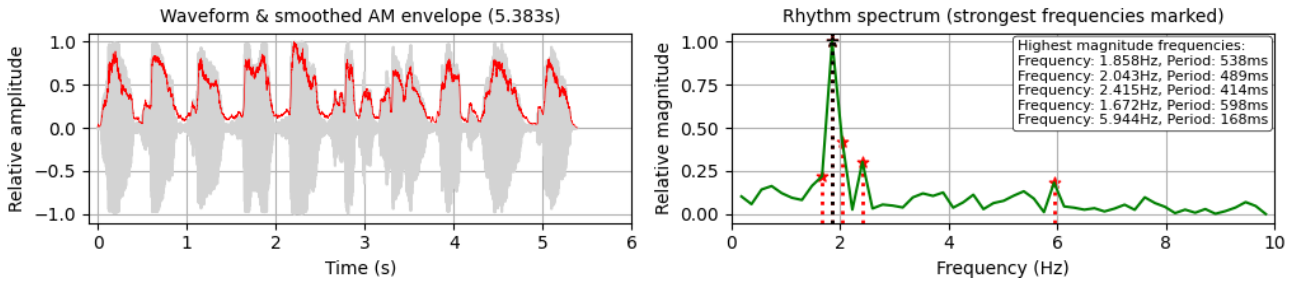

Figure 2: From Amplitude Envelope to Low Frequency Spectrum: British English, adult male.

The rhythmical sequence of ten numeral words can be clearly seen in Figure 2 (left). Two of the words are a little exceptional: *six* has a very short vowel flanked by voiceless consonants, and *seven* has two syllables, both with two short vowels. The other syllables are longer monosyllables. The frequency of the main rhythm, the numeral rate per second, is easy to calculate: 10 words divided by the duration 5.383 s yields 1.858 Hz, that is 1.858 words per second, average repetition period 538 ms.

By using spectral analysis with a Fast Fourier transform, more precise results can be calculated: Figure 2 (right) shows that the main frequency peak is indeed found at about 1.858 Hz, exactly as measured by hand, with other frequencies surrounding the main frequency, due to irregularities in syllable structure and durations. Higher, weaker frequency peaks, most clearly at 5.944 Hz (period of 168 ms) also appear, also due to syllable components and their durations.

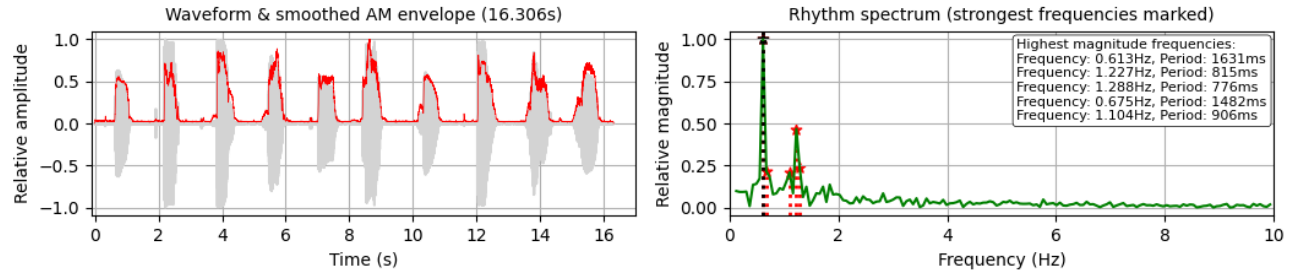

Figure 3: From Amplitude Envelope to Low Frequency Spectrum: Mandarin Chinese, adult female.

Figure 3 shows a rhythmical sequence of counting to ten in Mandarin Chinese: *yī èr sān sì wǔ liù qī bā jiǔ shí*. The four tones are marked with diacritics: 1, high flat tone (*yī*); 2, rising tone (*shí*); 3, fall-rise tone (*wǔ*); 4, falling tone (*èr*). The Mandarin speaker chose to speak with a much slower rhythm which peaks around 0.613 Hz, both measured and manually calculated, and further modulated by a binary syllable pattern at twice this frequency, around 1.227 Hz, showing that Mandarin has a simpler and more regular syllable structure than English, with an additional language-specific difference: the higher frequency peaks in English are absent in Mandarin, which only shows a noise pattern at these higher frequencies.

Much more can be said about speech rhythms. Research continues, not only on rhythms in sentences but also on open questions concerning the contribution of speech melody to speech rhythm, on the rhetorical rhythms which characterise longer speeches and lectures, and on the variation of rhythm during speaking (Gibbon and Li 2020). Parallel to investigations on physical properties of speech rhythm, communicative properties of rhythm in structuring and framing speech and in signalling attitudes and emotions are also being investigated (Couper-Kuhlen and Selting 2018) as well as interdisciplinary studies on the evolution of rhythm processing in speech and music and in human and animal communication (Kotz et al. 2018).

References

Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Asu, Eva-Liina and Francis Nolan. 2006. Estonian and English rhythm: a twodimensional quantification based on syllables and feet. In: *Proc. Speech Prosody* 2006.

Barbosa, Plínio A. 2002. Explaining cross-linguistic rhythmic variability via a coupled-oscillator model for rhythm production. In: *Proc. Speech Prosody* 1, 163–166.

Barbosa, Plínio A., M. Céu Viana, Isabel Trancoso. Cross-variety rhythm Typology in Portuguese. Proc. *Interspeech*, Brighton, 2009.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341–345.

Couper-Kuhlen, Elizabeth and Margret Selting. 2018. *Interactional Linguistics. Studying Language in Social Interaction*. Cambridge: Cambridge University Press.

Cummins, Fred and Robert Port. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26, 145–171.

Galves, Antonio, Jesus Garcia, Denise Duarte & Charlotte Galves. 2002. Sonority as a basis for rhythmic class discrimination. In: Bernard Bel & Isabel Marlien, eds. Proc. *Speech Prosody* 2002, Aix-en-Provence: Laboratoire Parole et Langage, 323–326.

Gibbon, Dafydd and Peng Li 2019. Quantifying and correlating rhythm formants in speech. *Proc. Linguistic Patterns in Spontaneous Speech*, Taipei, Academia Sinica.

Jassem, Wiktor 1952. *Intonation of Conversational English (Educated Southern British)*. Wrocław: Wrocławskie Towarzystwo Naukowe.

Jassem, Wiktor, David R. Hill and Ian H. Witten. 1984. Isochrony in English Speech: Its Statistical validity and linguistic relevance. In: Dafydd Gibbon and Helmut Richter eds. *Intonation, Accent and Rhythm. Studies in Discourse Phonology*. Berlin: Walther de Gruyter, pp. 203–225.

Jones, Daniel. 1909. *The Pronunciation of English*. Cambridge: Cambridge University Press.

Kotz, Sonja A., Andrea Ravignani and W. Tecumseh Fitch. 2018. The evolution of rhythm processing. *Special issue: Time in the Brain. Trends in Cognitive Sciences*, Vol. 22, No. 10, 896-910.

Low, Ee Ling, Esther Grabe and Francis Nolan. 2000. Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43:4, 377-401.

Palmer, Harold E. 1924. *English Intonation: With Systematic Exercises*. Cambridge: Heffer.

Pike, Kenneth L. 1946. *The Intonation of American English*. University of Michigan Publications: Linguistics: vol. 1.) Ann Arbor: University of Michigan Press.

Ramus, Franck, Marina Nespor and Jaques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292.

Roach, Peter. 1982. On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. In: Crystal, David, ed. *Linguistic Controversies: Essays in Linguistic Theory and Practice*. London: Edward Arnold, 73–79.

Tilsen Samuel and Keith Johnson. 2008. Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*. 124 (2): EL34–EL39. 2008. [PubMed: 18681499].

Todd, Neil P. McAngus and Guy J. Brown. 1994. A computational model of prosody perception. In: *Proc. the International Conference on Spoken Language Processing* (ICLSP-94), 127–130.