RHYTHM AND TIMING ACROSS LANGUAGES

Amalia Arvaniti

It is generally assumed that speech rhythm is based on timing that is, on durational relations between elements of the speech signal. Here I present research that questions the relation between timing and rhythm and propose a way forward that is of relevance to the cross-linguistic investigation of rhythm.

The connection between rhythm and timing is closely linked to the notions of rhythm class and isochrony which have been a feature of rhythm research for over 70 years. Despite the persistence of these views, supporting evidence has been hard to come by. Arvaniti (2012), based on data from eight speakers of English, German, Greek, Italian, Korean and Spanish, shows that the most recent conceptualization of isochrony, that quantified in rhythm metrics (e.g. Ramus et al. 1999; Grabe & Low 2002), is not accurate either: languages are not easily separated into classes based on metrics; metrics do not correlate with each other and are very sensitive to both inter-speaker variation and syllable complexity. Further, languages said to be in the same class do not share timing features captured by metrics: e.g., Naïve Bayes classifiers trained on English-Italian classified Korean data differently from classifiers trained on German-Spanish (Horton & Arvaniti ms.). Together, these results suggest that there is much more variability within each language and each rhythm class than previously assumed and that interval metrics do not transparently reflect these trends.

There is also evidence from perception that treating timing as the sole exponent of rhythm may not be ecologically valid. Arvaniti (2013) found that impoverished signals with only timing information, such as flat *sasasa*, do not elicit the same responses as richer signals that include segmental and F0 information, such as low-pass filtering. Arvaniti & Rodriquez (subm.) further find that discrimination results said to be due to rhythm class differences reflected in timing are actually due to differences in tempo between languages and, to a lesser extent, F0 patterns: once these differences are eliminated, discrimination becomes almost impossible. These results suggest that (a) discriminability may not be an appropriate measure of cross-linguistic differences in rhythm; (b) the relative timing of consonantal and vocalic intervals is not what creates percepts of rhythm.

Considered together these results suggest that the connection between rhythm and timing is less direct than hitherto assumed and that rhythm class distinctions are at best tenuous. Thus, in order to move forward we need to abandon the assumption that rhythm is based exclusively on timing. Following Arvaniti (2009) it is proposed that *timing* be used to denote all aspects of the durational organization of speech and not as a synonym for *rhythm*. In turn, *rhythm* is defined as the perception of series of stimuli as series of groups of similar and repetitive pattern (cf. Clarke, 1999). The question that arises then is what these stimuli and groupings might be in speech if they are not due exclusively to timing.

A plausible answer comes from recent results using envelope-based metrics (henceforth *EMs*; Tilsen & Arvaniti subm.). EMs capture periodicities in speech by extracting frequency components (Intrinsic Mode Functions or IMFs) from the spectral envelope. IMFs are independent of human segmentation into vowels and consonants yet correspond well to syllabic (IMF_1) and supra-syllabic (IMF_2) elements in speech. The application of EMs to the corpus of Arvaniti (2012) shows interesting parallels with interval metrics. At the same time, however, EMs suggest that cross-linguistic differences in periodicity are not substantial: syllable-level periodicities

appear at similar and stable frequencies across languages. Crucially, supra-syllabic periodicities appear at frequencies clustering around 2.5 Hz, indicating the presence of a supra-syllabic prominence every 400-500 ms. This frequency is close to *natural tempo* (Clarke 1999) and the rate of stress occurrence reported by Dauer (1983) for English, Thai, Spanish, Italian and Greek; it is also within the Delta band (1.5–4 Hz) of oscillating neuronal networks crucial for speech processing and rhythm entrainment (e.g., Luo and Poeppel 2007).

The EM results indicate that rhythm is based on general principles of speech organization that go beyond cross-linguistic differences in prosodic structure; e.g. results are not dramatically different between (i) stress-timed English and German, (ii) Greek, Italian and Spanish, which are said to be syllable-timed but have stress, and (iii) Korean which is said to be syllable-timed but does not have stress. They further suggest that cross-linguistically syllables can be seen as the basic series of stimuli that are grouped into patterns by the presence of prominences repeated every half a second or so. Regarding Korean in particular these results are corroborated by recent research on cycling (Chung & Arvaniti 2012): data from ten speakers show that accentual phrase onsets (henceforth *ap*s) exhibit stable phasing in a simple cycling task, while *ap* durations agree well with the mean instantaneous frequency of IMF_2 for Korean. In turn these results support the idea that supra-syllabic periodicities in the speech signal need not be due to stress and feet as conventionally understood, but also appear in languages in which stress and foot structure are not present.

In conclusion, this body of research suggests that we need to consider alternatives beyond timing in order to understand patterns of periodicity in speech. One of these alternatives, envelope-based metrics, is particularly suited for cross-linguistic research as it provides an objective way of investigating rhythm. Nevertheless, the role of native input should not be underestimated or ignored: native intuitions should guide researchers to ask language-appropriate questions about rhythm which, in turn, should lead to a better understanding of both commonalities and differences in rhythm across languages.

References

Arvaniti, A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66: 46-63.

Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *J.Phon.* 40: 351-73.

Arvaniti, A. 2013. Rhythm classes and speech perception. In O. Niebuhr (ed.) *Understanding Prosody – The Role of Context, Function, and Communication*. Walter de Gruyter.

Arvaniti, A. & T. Rodriquez. Subm. The role of rhythm class, tempo and F_0 in language discrimination: implications for the nature of speech rhythm.

Chung, Y. & A. Arvaniti. 2012. Speech cycling in Korean. To be presented at the "Port Workshop: Perspectives on Rhythm and Timing," Glasgow, UK, July 19-21, 2012.

Clarke, E. F. 1999. Rhythm and timing in music. In D. Deutsch (ed.) *The Psychology of Music*. New York: Academic Press, 473-500.

Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *J. Phon*. 11: 51-62.

Grabe, E. & E.-L. Low. 2002. Acoustic correlates of rhythm class. In C. Gussenhoven & N. Warner (eds) *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 515-46.

Horton, R. & A. Arvaniti. 2012. Clustering and classification of the rhythm metrics. UCSD ms.

Luo, H. A. Poeppel. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54: 1001-10.

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

Tilsen, S. & A. Arvaniti. Subm. Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages.