

Dynamic Time Warping and Prosodic Prominence

Valentina Schettino (University of Naples “L’Orientale”), **Antonio Origlia** (University of Naples “Federico II”), **Francesco Cutugno** (University of Naples “Federico II”)
vschettino@unior.it, antori@gmail.com, cutugno@unina.it

One of the first issues to be solved when approaching the annotation of prosodic prominence (cf. Terken, 1991) is the kind of rating scale to use for prominence levels. In the literature, different kind of scales have been used. Lately, a new methodology has been developed (cf. Samlowski & Wagner, 2016), which exploits the prosody-gesture link; prominence perception is related to the beating movement: participants of the experiment are asked to listen to some short sentences and reiterate them by beating on a DrumPad, modulating the intensity of the beat in a directly proportional way. The drumming task permits an easy, intuitive processing of prosodic prominence, allowing raters to produce a fine-grained annotation without necessarily being experts; moreover, this procedure proved to be very fast, thus enabling the annotation of very large corpora in a consistent way. The main problem with this approach is that the drumming procedure leaves raters free to decide how many beats they hear. Traditionally, instead, manually annotated corpora for prosodic prominence have used a strict correspondence of syllables and annotation units (cf. Fant & Kruckenberg, 1989; Wightman 1993; Streefkerk et al. 1999; Jensen 2004). The drumming procedure has the advantage of removing this constraint; this choice, however, implies that the drumming associated with a specific file may not contain an amount of beats that is equal to the number of reference syllables. In Samlowski & Wagner (2016), authors examined only drummed sentences in which the number of expected syllables and the number of beats coincide, as their aim consists mainly in the validation of the drumming procedure; on the contrary, we intend to develop a system that can allow to use most of the drummed sentences. To reach this aim, we develop an objective alignment process using the Dynamic Time Warping (DTW, cf. Sakoe & Chiba, 1978). This procedure suits our aim for different reasons. Firstly, the best alignment between the two sequences, the so-called *warping path*, does not result from a subjective choice, but is the manifestation of objective parameters. The second advantage of the DTW approach is that it allows us to evaluate the quality of the annotations: the minimum distance provided by the DTW algorithm – interpreted as the effort that the algorithm has to do in order to connect beats and reference syllables – gives indications in this sense. Furthermore, we can calculate the qualitative threshold for each rater, leaving aside all the drummed files that diverge too much from the reference. Little discrepancies between the amount of drummed beats and syllabic units are acceptable for our analysis; indeed, we regard them as possible expressions of perceptual processes. On the contrary, drumming sequences that greatly differ from the reference ones cannot be used for further analyses and must be discarded.

Summing up, we think that this procedure can ameliorate the DrumPad method, which in turn could result in a better annotation system, applicable to large corpora of spontaneous speech.

References

- Fant, G., & Kruckenberg, A. (1989). Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR*, 2(1989), 1-83.
- Jensen, C. (2004). Stress and accent. Copenhagen: University of Copenhagen. Ph.D. dissertation.
- Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1) pp. 43– 49.
- Samlowski, B., & Wagner, P. (2016). PromDrum-Exploiting the prosody-gesture link for intuitive, fast and fine-grained prominence annotation. *Proceedings of Speech Prosody 2016*.

- Streefkerk, B. M. (2002). *Prominence. Acoustic and lexical/syntactic correlates*. Doctoral Thesis. Utrecht: LOT.
- Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *The Journal of the Acoustical Society of America*, 89(4), 1768-1776.
- Wightman, C. W. (2002). ToBI or not ToBI?. In *Speech Prosody 2002, International Conference*.