Prosodic Phrasing and Syllable Prominence in Spoken Prose.
A Validated Coding Manual

Isabelle Franz[1,2], Christine A. Knoop[1], Gerrit Kentner[1,2], Sascha Rothbart[3], Vanessa Kegel[1], Julia Vasilieva[4], Sanja Methner[1], Mathias Scharinger[5], and Winfried Menninghaus[1]

[1]Max Planck Institute for Empirical Aesthetics, Frankfurt; [2]Goethe University Frankfurt; [3]Univerity of Bonn; [4]University of Côte d'Azur; [5]Philipps University Marburg

Metrical grids are supposed to reflect relative syllable prominence (Liberman & Prince, 1977), and partly account for the domains of the Prosodic Hierarchy (Halle & Vergnaud, 1987). However, their use for empirical studies is limited to highly controlled and short sentences. Also, current systems using metrical grids for syllable prominence prediction focus on decoding small verses (for poetry see Lerdahl, 2001), or on syntax/semantic-based automatic decoding of sentences that need to be annotated syntactically (Windmann et al., 2011). A replicable system for manually coding syllable prominence and prosodic boundaries in longer sentences or even texts is lacking so far, let alone its validation with the phonetic realization.

Based on work in the fields of metrical phonology (Kiparsky, 1966; Liberman & Prince, 1977) and existing prominence and pause coding systems (Gee & Grosjean, 1983; Windmann et al., 2011), we developed a manual for coding syllable prominence (yielding up to 9 degrees of prominence) and prosodic boundaries (with 6 degrees of juncture, including positions where a boundary is particularly unlikely). The manual consists of a set of rules that are to be applied in a prescribed order; these rules mainly refer to simple cues in the text, like word/syllable count, part of speech, word position and punctuation. The coding system is based on the understanding that syllable prominence and boundary strength determine each other (Franz et al., 2022).

Three independent annotators applied the coding system to the beginning pages of four different German novels (~90 000 syllables, Fleiss kappa .88). For the phonetic validation, eight professional speakers read the texts aloud. We annotated the speech signal automatically with MAUS (Schiel, 1999). Using PRAAT (Boersma & Weenink, 2019), we extracted F0 range and duration for each syllable and compared it to predicted syllable prominence (Figure 1-2). We further compared pause duration to predicted prosodic boundary strength (Figure 3). The validation with the speech signal shows that our annotation system reliably predicts syllable prominence and prosodic boundaries. In comparison to Gee and Grosjean (1983), who developed a system to predict pauses from text with an infinite number of boundary degrees, our system generates six degrees. This is comparable to GTOBI (Baumann et al., 2000) where the speech signal is annotated. Since our annotation works with plain text, there are additional potential applications of the coding system, covering synthetic speech and (psycho)linguistic research on prosody.

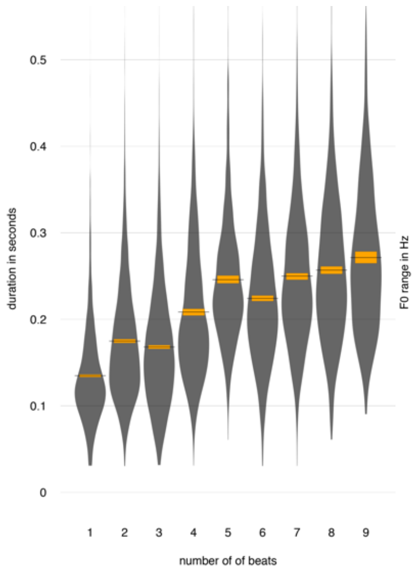

Figure 1: Syllable duration (in sec) by predicted syllable prominence (number of beats). The yellow bar shows the CI for the mean.

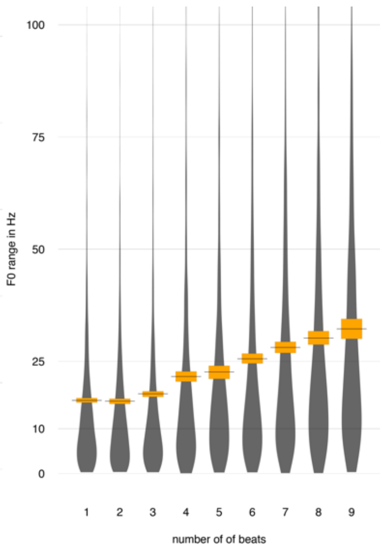

Figure 2: F0 range (in Hz) by predicted syllable prominence (number of beats). The yellow bar shows the CI for the mean.

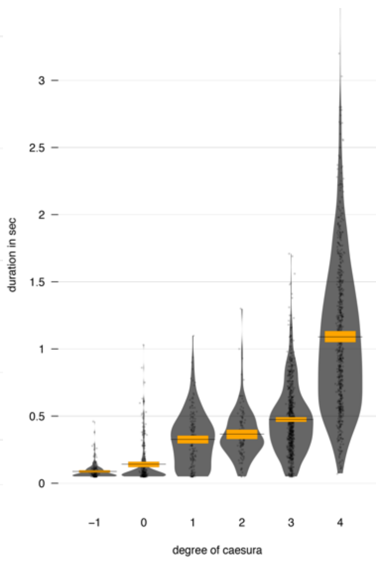

Figure 3: Pause duration (in sec) by predicted strength of prosodic boundaries (scale from -1 to 4). The yellow bar shows the CI for the mean.

References

Baumann, S., Grice, M., & Benzmüller, R. (2000). GToBI-a phonological system for the transcription of German intonation. In *Prosody* (pp. 21-28).

Boersma, P., & D. Weenink (2019). *Praat: doing phonetics by computer* (Version 6.0. 52)[Windows].

Franz, I., Knoop, C. A., Kentner, G., Rothbart, S., Kegel, V., Vasilieva, J., Methner, S., Scharinger, M. & Menninghaus, W. (2022). Prosodic Phrasing and Syllable Prominence in Spoken Prose. A Validated Coding Manual. OSF Preprint.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive psychology*, *15*(4), 411-458.

Halle, M., & Vergnaud, J. R. (1987). Stress and the cycle. *Linguistic Inquiry*, *18*(1), 45-84.

Kiparsky, P. (1966). Über den deutschen Akzent. Studia grammatica, 7, 69-98.

Lerdahl, F. (2001). The sounds of poetry viewed as music. *Annals of the New York Academy of Sciences*, *930*(1),337–354.

Liberman, M., & A. Prince (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2): 249–336.

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. *14th International Conference of Phonetic Science*, 1–7 August 1999, San Francisco.

Windmann, A., I. Jauk, F. Tamburini, & P. Wagner. (2011). Prominence- based prosody prediction for unit selection speech synthesis. *12th Annual Conference of the International Speech Communication Association*, 27–31 August 2011, Florence.