

Pause duration and other prosodic boundary cues are not monotonically correlated
Gerrit Kentner, Isabelle Franz, Christine Knoop, Winfried Menninghaus

Research on the phonetic expression of prosodic boundaries suggests phrase-final lengthening, pitch excursion and pause duration to be positively correlated, i.e. the stronger the prosodic break, the longer the duration of phrase-final syllables, the greater the pitch excursion, and the more likely or longer a pause is predicted to be (Cooper & Paccia-Cooper 1980, Wightman et al 1992, Krivokapic 2007, among others). However, most studies we are aware of are concerned with only a single boundary cue (e.g. final lengthening or pooling of final lengthening plus pause duration) and/or with only a limited set of prosodic boundary strengths (e.g. comparing only two levels). Exploring a large sample of read prose, we consider pause duration, phrase-final lengthening, and pitch excursion as potential boundary cues and compare these across five predicted levels of boundary strength. The results show that these phonetic cues are not correlated with boundary strength in a simple fashion. We suggest this to mean that pause duration on the one hand, and final lengthening and pitch excursion on the other, have different functional imports in speech production.

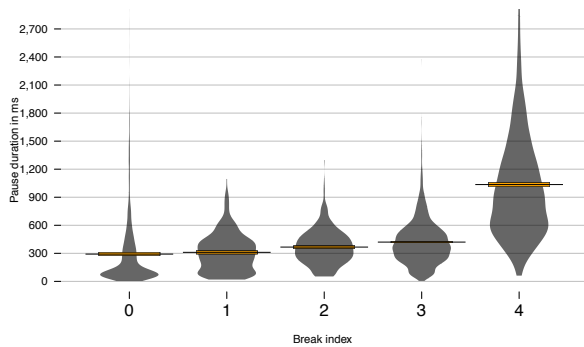
Methods: We recorded eight professional speakers each reading aloud four prose text samples (~1500-1800 words). The read texts (with a total of >50,000 spoken words, ~6h of speech) were automatically segmented for words and syllables using MAUS (Schiel et al. 1999). Using praat, we extracted duration, pitch, and intensity values for each syllable.

We applied a manual coding scheme (Franz et al., forthcoming) to predict prosodic breaks on the basis of textual features. This coding system presupposes 5 degrees of boundary strength (0: no break predicted; 1: phrase break, no comma; 2: short comma phrase; 3: long comma phrase; 4: sentence boundary).

Results: The plots show a) pause length broken down by predicted phrase break, b) duration of the syllable directly preceding the break, and c) pitch excursion on the final syllable. While pause length monotonically increases with predicted boundary strength, this is not the case for the other boundary cues. Compared to the no-boundary condition (break index 0), pre-boundary syllables are lengthened, but final lengthening is strongest for break index 2 and decreases through indices 3 and 4. Similarly, pitch excursion is increased on pre-boundary syllables but this increase peaks at break index 2 and decreases through indices 3 and 4.

Discussion: The monotonic increase of pause duration along the predicted scale, and the non-monotonic increase of the other prosodic boundary cues indicates that these phonetic signals reflect different processes in speech production. The increasing pause durations probably reflect the closure of units of increasing size and planning of upcoming clauses (Ferreira 1991). We tentatively suggest that the phonetic cues on the pre-boundary syllables reflect current planning complexity: Breaks provide a time window for speech planning, and planning complexity is high as long as clauses and sentences are not finalized. Breaks with commas (index 2 and 3) offer more planning time than breaks without comma (index 1). However, index 3 breaks are more likely to close off a clause, while at index 2, planning for the current clause is more likely to be still ongoing. Therefore, planning complexity at break index 2 will be higher than at break index 3. Finalized clauses (break index 3) and sentences (index 4) require less or no time for current phrase planning, with syllable duration and, concomitantly, pitch excursion consequently decreasing.

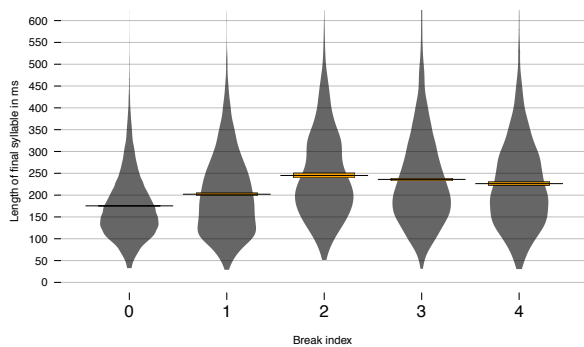
In sum, this research shows that the suggested monotonic correlation between pause duration and other phonetic boundary cues is not valid.



Successive differences contrasts between break indices (negative differences in bold)

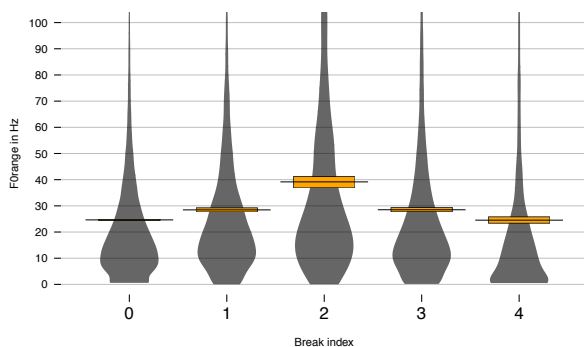
Pause duration (log):

1-0: $t = 5.012$, $p < 0.001$
 2-1: $t = 35.996$, $p < 0.001$
 3-2: $t = 6.214$, $p < 0.001$
 4-3: $t = 75.028$, $p < 0.001$



Duration of pre-boundary syllable (log):

1-0: $t = 15.67$, $p < 0.001$
 2-1: $t = 13.83$, $p < 0.001$
3-2: $t = -3.24$, $p < 0.001$
4-3: $t = -4.62$, $p < 0.001$



F0 range on pre-boundary syllable (log):

1-0: $t = 13.46$, $p < 0.001$
 2-1: $t = 7.11$, $p < 0.001$
3-2: $t = -8.34$, $p < 0.001$
4-3: $t = -11.27$, $p < 0.001$

References

- Cooper & Paccia-Cooper (1980). *Syntax and speech*. Harvard University Press.
- Ferreira F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *J. Mem. Lang.* 30, 210 – 233.
- Franz, I. et al. (forthcoming). Syllable prominence and prosodic phrasing in spoken prose – a manual for their prediction from text. Ms MPI for Empirical Aesthetics, Frankfurt.
- Krivokapic J. (2007). The planning, production, and perception of prosodic structure. PhD thesis, USC Los Angeles.
- Schiel, F. (1999). Automatic phonetic transcription of nonprompted speech. *Proc. ICPHS*, 607–610, San Francisco, USA.
- Wightman et al. (1992) Segmental durations in the vicinity of prosodic phrase boundaries. *JASA* 91.3, 1707-1717.