Is prosodic phrase structure planned? Evidence from phrasal lengthening, autocorrelation, and Markov statistics in spontaneous speech

Sam Tilsen

Cornell University

A wealth of experimental evidence shows that prosodic phrase structure correlates with phonetic variation, and so it is tempting to infer that utterance plans are organized into a hierarchy of prosodic units. This inference seems to make sense for relatively long stretches of speech elicited in laboratory contexts, but what about spontaneous conversational speech? If utterance plans are phrasally organized prior to utterance initiation, we would expect the following to hold: (i) *phrase-final lengthening* should be well predicted by transcribed boundary strength or phrase type; (ii) *negative correlation* should be observed between adjacent phrase sizes and durations, as is often the case for other types of units in the prosodic hierarchy; and (iii) *multi-phrase stretches* of fluent speech should be common enough to warrant the need for prosodic phrase structure in the first place. These predictions were not supported by analyses of conversations in Switchboard NXT corpus. Instead, observed patterns may suggest that conceptual-syntactic planning and lexical retrieval are more influential than prosodic phrase structure in the planning and production of spontaneous speech.

Method: NXT Switchboard (Calhoun et al. 2010) is an extensively annotated subset of the Switchboard corpus of telephone conversations, totaling 642 conversations and about 830,000 words. 45 conversations include ToBI annotations with break indices and major/minor phrase labels. For prediction (i), different types of predictor variables were compared in their ability to account for variance in phrase-final phone duration (residualized over phone identity and syllable stress). Predictor types considered were: break index of the following boundary, phrase type (major/minor), and number of words/phrases in the following stretch of speech. For (ii), multi-phrase sequences were extracted, and lag-1 autocorrelation was calculated for phrase duration, number of words and number of syllables within phrase. For (iii), Markov chain transition probabilities were estimated with states of: word, (silent) pause, filled pause, and discourse marker (e.g. *well*, *okay*, *yes*, etc.).

Results: Regarding (i) *phrase-final lengthening*, it was found that ToBI break indices and phrase categories were relatively poor predictors of residualized phrase-final phone duration, compared to other predictors such as number of phrases/words in following fluent speech. This can be seen in the boxplots in Fig. 1; for example, the number of fluent words after a phrase boundary (rightmost panel) accounts for nearly four times as much variance in phrase-final lengthening as phrase category (leftmost panel). Regarding (ii) *negative correlation*, measures of phrase size (i.e. duration, number of words/syllables) in multi-phrase sequences did not exhibit negative lag-1 autocorrelation (see Fig 2). Note that negative autocorrelation is expected for other types of subunits in the prosodic hierarchy, such as syllables within feet, or segments within syllables. To the contrary, weak positive lag-1 autocorrelations were observed for the number of words in a phrase and phrase duration. These positive autocorrelations are possibly due to interspeaker variation in fluency or speech rate. Regarding (iii) *multi phrase stretches*, these may not be so common: fewer than 7% of fluent stretches of speech consisted of more than 15 words; indeed, Markov chain transition probabilities in Fig. 3A show that the probability of transition to a non-word state after a word is around 0.16 (Fig. 3D), despite the word state having a nearly 80% occupancy (Fig. 3B). The rarity of long stretches of fluent speech questions whether speakers need to prepare a phrasal structure in most utterances.

Conclusion: None of the three predictions of the hypothesis that speakers prepare a prosodic phrase structure were supported in analyses of spontaneous conversational speech. This may suggest that prosodic phrasal structure is only planned sometimes, or that factors such as lexical retrieval and conceptual-syntactic planning are more influential than prosodic phrase structure in speech production.

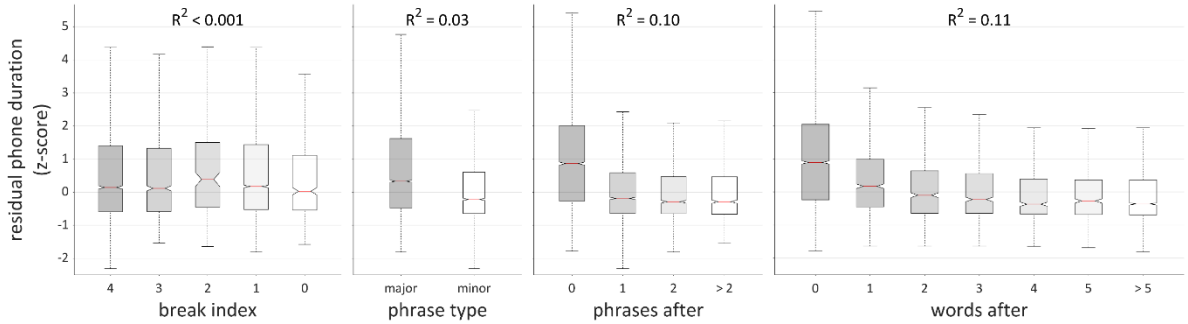

Fig. 1. Boxplots of residual phone duration (z-scored) for phrase-final phones, grouped by four different types of predictors: ToBI break indices, prosodic phrase category, number of following phrases after, and number of following words.

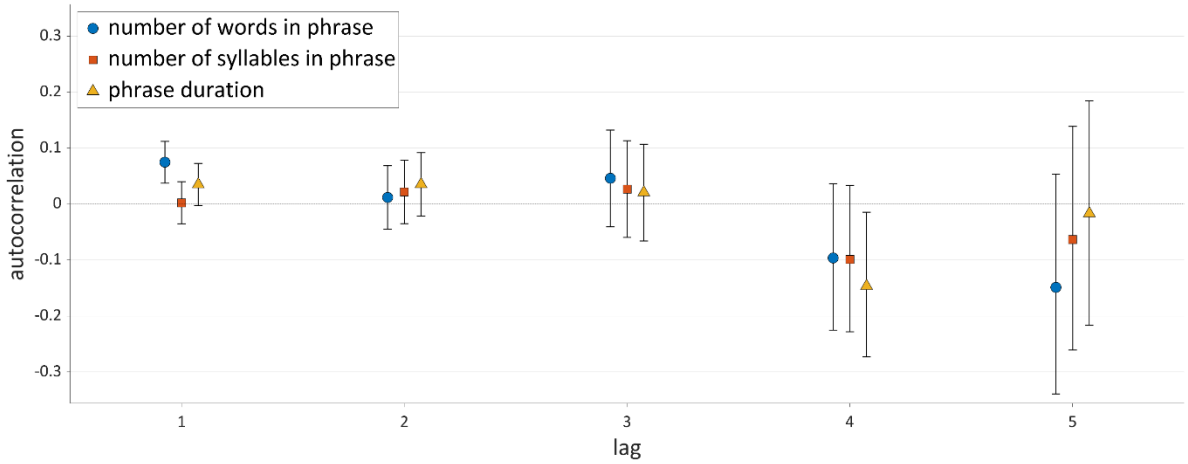

Fig. 2. Autocorrelations (lags 1-5) of duration, number of words, and number of syllables for phrases from fluent stretches of speech. 95% confidence intervals are shown.

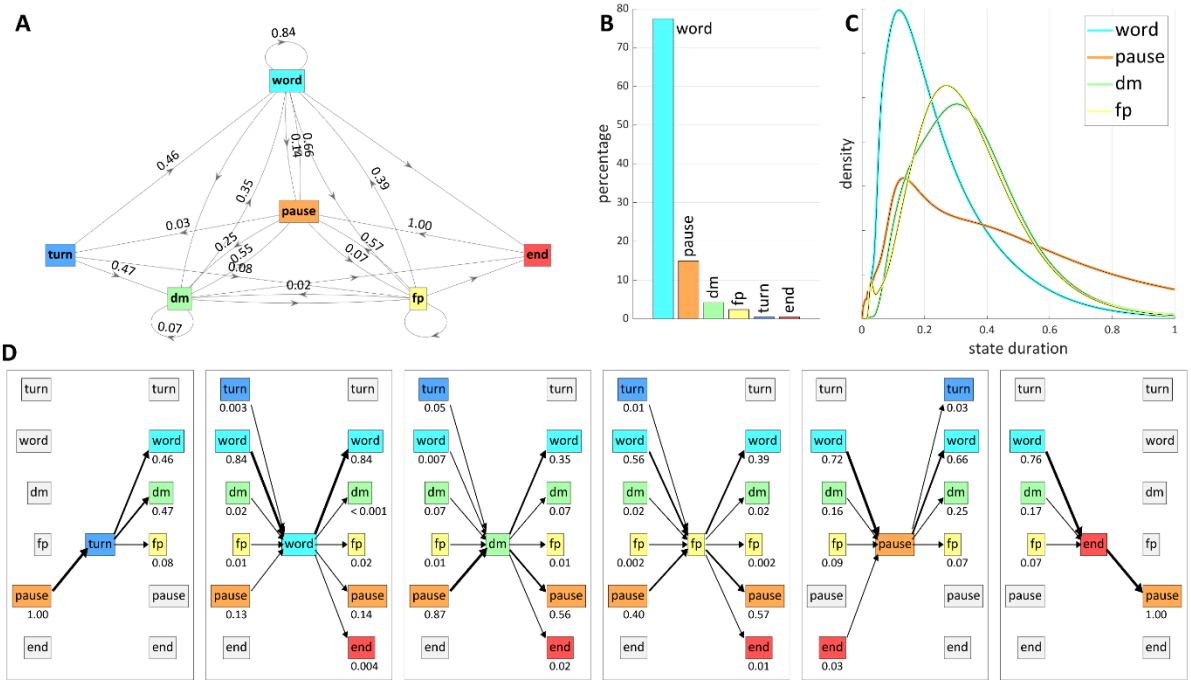

Fig. 3. (A) Markov chain model of speech states from the Switchboard NXT corpus: turn start, word, pause, discourse marker (dm), filled pause (fp), and turn end. For clarity transitions < 0.02 are not labeled. (B) State occurrence percentages. (C) State duration densities. (D) Forward and backward transition probabilities by state.