

Is the intermediate phrase the basic unit of speech production planning? Evidence from individual differences

Jason Bishop & Darlene Intlekofer

A range of phonetic and phonological patterns indicates that speech production planning unfolds in relatively large chunks—chunks that may be better defined in terms of phrase-level prosodic units than in terms of one- or two-word sequences (Keating & Shattuck-Hufnagel, 2002). Of recent interest has been the question of planning's flexibility—that is, the extent to which planning scope is sensitive to the various demands facing the speaker (Wagner et al., 2010; Krivokapić 2012). Notably, Levelt (1989) and others have assumed that phonological planning in part utilizes working memory resources, and thus working memory capacity (WMC) may represent a speaker-specific constraint on planning scope (Wagner et al., 2010). The present study tested the hypothesis that the extent of speakers' planning is positively related to their WMC, exploring two prosodic variables associated with planning scope: *phrase length* and *pre-phrasal pause duration*.

With respect to *phrase length*, we predicted that speakers with lower WMC, if they engage in shorter-range planning, would tend to produce shorter prosodic phrases (intonationally-defined within the AM model for English in Beckman & Pierrehumbert, 1986), a prediction that has some support in the literature on implicit prosody (Swets et al., 2007). Notably, however, it is not clear a priori how speakers might accomplish this. For example, for the same string of words, a speaker with lower WMC could shorten phrases by either (a) parsing the string into a greater number of shorter Intonational Phrases (IP), or (b) by parsing it into the same number of IPs that instead contain a greater number of shorter intermediate phrases (ip).

With respect to *pre-phrasal pause durations*, we predicted that the interactive effects of upcoming phrase length and upcoming phrase structure found by Krivokapić (2007/2012) would be further mediated by WMC (see also Petrone et al., 2011). When upcoming phrases are longer (in syllables count), preceding pause durations are longer as well, reflecting the time taken to plan that material (Whalen & Kinsella-Shaw, 1997). However, this effect is less apparent when the upcoming stretch of speech is parsed into more than one prosodic unit, and this can be explained as follows. When the upcoming stretch of speech contains only one prosodic unit, the speaker must begin planning that entire unit, and so its overall length greatly influences planning time before execution. However, if the stretch of speech is divided into two prosodic sub-units, the speaker has the option of planning only the first sub-unit during the pause, and then (more incrementally) planning the second sub-unit while producing the first—reducing the relevance of the stretch of speech's overall length to predicting pause duration. Our prediction was that, other things being equal, speakers with lower WMC should tend to prefer the more incremental strategy, and thus produce *shorter* pauses when an upcoming IP contained two or more ips.

These predictions were tested on a corpus of read speech collected in the context of a previous production study with 100 native speakers of American English. Speakers read aloud the same 160-word passage and completed the reading span task (Daneman & Carpenter, 1980). The phrase structure of the produced passages was identified using ToBI conventions for Mainstream American English (Beckmann & Ayers Elam, 1997) as described in Bishop & Kim (2018). In brief, two annotators determined the presence of all ip-level boundaries (ToBI labels of 3-, 3, or 3?) and all IP-level boundaries (4-, 4, or 4?); agreement between the two regarding the presence of a break above the word level was 92.5% ($\kappa = .76$) and 100% of the disagreements concerned the presence of an ip boundary versus the absence of a boundary. Where both annotators agreed a boundary above the word level was present, agreement as to its status as ip- vs. IP-level was 94.5% ($\kappa = .88$). Because the analysis required definitive

decisions about the locations and sizes of boundaries for all speakers, and because we wished to only analyze boundaries about which we had the most certainty, the relatively small proportion of disagreements were settled in the direction of the annotator who marked a smaller degree of juncture. The length (in syllables) of all speakers' fluent ips and IPs were then counted, and the durations of all pauses preceding fluent IPs were measured.

Results showed the following. First, WMC was predictive of speakers' phrase lengths, but only significantly so when defined as ips, and WMC did not predict which of the strategies described above that speakers chose. Similarly, WMC was found to predict pre-phrasal pause durations when an upcoming IP was grouped into smaller ips, but not when an upcoming utterance consisted of multiple IPs. We discuss the implications of these patterns, which suggest a special prominence for the intermediate phrase in speech planning, possibly similar to the prominence it has been claimed to have in sentence processing (e.g. Schafer's 1997 Prosodic Visibility Hypothesis).

References

- Beckman, M. & Pierrehumbert, J. (1986) Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.
- Beckmann, M., & Ayers Elam, G. (1997). Guidelines for ToBI labeling (Version 3). Unpublished ms. Ohio State University.
- Bishop, J. & Kim, B. (2018). Anticipatory shortening: Articulation rate, phrase length, and lookahead in speech production. *Proceedings of Speech Prosody* 9, 235–239.
- Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research*, 20(6), 445–464.
- Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 101, 112–156.
- Krivokapić, J. (2007). The planning, production, and perception of prosodic structure. Ph.D. dissertation, University of Southern California.
- Krivokapić, J. (2012). Prosodic planning in speech production. In S. Fuchs, M. Wehrich, D. Pape, & P. Perrier (eds.): *Speech planning and dynamics*. Peter Lang. pp. 157–190.
- Levelt, W.J. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- Petrone, C., Fuchs, S., & Krivokapić, J. (2011). Consequences of working memory differences and phrasal length on pause duration and fundamental frequency. In Y. Laprie and I. Steinder (Eds.): *Proceedings of the 9th International Seminar on Speech Production*. (pp. 393–400).
- Schafer, A.J. (1997). Prosodic Parsing: The Role of Prosody in Sentence Comprehension. Ph.D. dissertation, University of Massachusetts.
- Swets, B., Desmet, T., Hambrick, D., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: a psychometric approach. *J. Experiment Psychology: General*, 136, 64-81.
- Wagner, V., Jescheniak, J., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 423–440.
- Whalen, D., & J. Kinsella-Shaw (1997). Exploring the relation of inspiration duration to utterance duration. *Phonetica* 54, 138–152.