10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Lower Working Memory Capacity is Associated with Shorter Prosodic Phrases: Implications for Speech Production Planning

Jason Bishop, Darlene Intlekofer

City University of New York
jbishop@gc.cuny.edu, dintlekofer@gradcenter.cuny.edu

Abstract

The present study investigated speech production planning from an individual differences perspective. In particular, we explored the possibility that cross-speaker variation in prosodic phrase length—assumed to reflect, in part, variation in speakers' planning scope—is systematically related to individual differences in working memory capacity—a cognitive resource that early phonological planning may utilize. Connected speech from a read passage, produced by 100 American English speakers, was analyzed for phrase structure, defined within the Autosegmental-Metrical framework, and the lengths of speakers' intermediate phrases and Intonational Phrases were calculated. Results showed that shorter reading spans (a measure of verbal working memory) were associated with shorter spoken phrase lengths, significantly so in the case of intermediate phrases. The basic findings lend support to the idea that planning is to some extent flexibly dependent on internal and external pressures facing the speaker. We discuss the implications of these findings for models of speech production and models of prosodic interfaces.

Index Terms: phrasing, planning, individual differences

1. Introduction

1.1. Prosody's role in planning utterances

According to the "prosody first" view of speech production planning ([1]), the earliest stages of phonological planning involve the creation of the upcoming utterance's phrase-level prosodic structure, and thus speakers are assumed to generally plan in multi-word chunks ([2]). Among the evidence that speakers engage in such extensive lookahead are phonetic effects that depend on overall phrase or utterance length (for other types of evidence, see [1] and [3]). For example, it has been found that when a prosodic phrase is longer (in units such as words or syllables):

- a preceding silent pause is longer ([4] – [9])
- the segments and syllables throughout the phrase tend to be shorter ([10] – [13])
- phrase-initial F0 peaks tend to be higher ([8],[14],[15])

In addition to being dependent on phrase length, these phonetic adjustments have also been shown to be dependent on phrase structure, as they are found more consistently in simple phrases (i.e. Intonational Phrases that consist of a single lower-level intermediate phrase; ([7],[8],[16]). Taken together, such patterns suggest speakers have advanced knowledge that stretches as far as an entire Intonational Phrase or utterance—hard to accommodate in models that assume planning unfolds in word-sized units ([17],[18]) rather than phrase-sized units.

At the same time, an idea that has been of increasing interest in recent years is that planning may be to some extent flexibly dependent on pressures facing speakers ([19],[20],[16]; see also [21],[22]). That is, rather than adhering rigidly to a unit of a particular size, speakers plan in larger or smaller chunks depending on pressures that can be construed as either *external* (e.g. properties inherent to the speaking conditions, such as the need to speak more quickly or more clearly) or *internal* (e.g. limitations on the cognitive resources that planning requires).

Of particular interest to us here is an internal pressure, namely limitations on working memory. Working memory has long been assumed important to the encoding of lexical and syntactic information during speech production ([17],[18][23],[24]). And while its role in the encoding of sound-based levels of representation is far less-well studied (see [25] for review), recent findings based on sentence initiation times suggest that speakers with lower working memory capacity tend to plan in smaller chunks at one or more levels of representation (e.g. [26],[27]). It is therefore plausible that the scope of speakers' phonological planning may be similarly constrained by working memory resources (e.g. [14],[28],[29]).

The idea of flexible planning raises interesting questions for the prosody-first view, where early phonological planning is assumed to unfold relatively non-incrementally. For example, in a language like English ([30]), what would it mean for planning units to be prosodic in nature, yet flexible? In particular, it is difficult to see how speakers could be said to plan in units lower in structure than Intonational Phrases, given that the timing effects mentioned above suggest that English speakers have advanced knowledge spanning at least this unit (see especially the correlations in [13]).

One possibility is that speakers, in challenging circumstances, produce a very rough sketch of the entire Intonational Phrase (as suggested by [1]), but in a second stage of (still phonological) planning, shift their attention to a lower level of structure. This is our interpretation of the strategy described previously by [16]. If speakers in fact engage in this sort of strategy, we might expect it to be apparent in their preference for certain phrase structures when speaking under pressure, such as Intonational Phrases that contain (and thus can be partitioned into) more than one intermediate phrase. In this case we would expect the speaker's Intonational Phrase length to be relatively unaffected, but for the greater number of intermediate phrases produced to be on average shorter.

Another possibility is that speakers, rather than prioritizing higher or lower levels of structure in this way, simply parse less material into their Intonational Phrases to begin with, making little adjustment to their structure. In this scenario, variation in the scope of speakers' planning may correlate with the length of their Intonational Phrases, but not necessarily their internal structure, and thus also not the average lengths of their constituent intermediate phrases.

In summary, then, we derive two basic predictions when we attempt to accommodate flexibility in a prosody-first model of planning, which seems to require a more length-based than unit-based definition of "flexible". First, planning considerations should leave their mark on speakers' phrasing choices—i.e. the

 10.21437/SpeechProsody.2020-39

complexity of their structures. Second, when facing pressures that limit planning scope, the choices they make should result in shorter phrases at some level of prosodic structure.

1.2. Working memory capacity & spoken phrase length

Motivated by the points made above, the present study aimed to explore the relationship between individual differences in speakers' working memory capacity (WMC) and prosodic phrasing. The target language was American English, and we define phrasing phonologically, based on the AM model proposed in [30]. We assumed that, to the extent that prosodic constituents reflect planning units, speakers with lower WMC—an internal constraint on planning scope—should tend to parse utterance material in such a way as to produce shorter phrasal constituents at some level. Our basic prediction, based on the discussion above, was that speakers with lower WMC will utilize one or both of the two strategies on the right in Fig.1 more often than higher WMC speakers, who should tend to favor the more monolithic structure on the left. Other things being equal, these kinds of differences in phrasing will be detectable in simple measures of phrase length; parsing the same amount of utterance material into a greater number of Intonational Phrases (henceforth IP) necessarily means shorter IPs on average, and parsing an IP into a greater number of intermediate phrases (ip) means shorter ips on average.

Before going on to the production study that tested this, we note that a positive correlation between WMC and prosodic phrase length has been assumed for some time in the literature on implicit prosody (i.e. subvocal prosody generated during silent reading; [32],[33]), and has been the basis explaining certain sentence comprehension patterns ([31],[34]). Despite this, we know of only one study investigating the role of WMC on the phrasing of overtly produced speech.

In that study, [35] measured prosodic phrase lengths (in their case defined as inter-pause intervals/utterances) produced by native English speakers whose WMC was manipulated experimentally. The authors used a dual-task paradigm that required speakers to complete a high-WMC load distractor task while simultaneously producing memorized sentences. Interestingly, the authors found the opposite pattern that we predict; when speakers were multi-tasking in their study (and thus should have had reduced WMC), they produced longer, not shorter, phrases. Under these conditions, however [35]'s speakers also produced faster speech. One possibility is that these speakers increased their speech rate as part of a strategy specific to the challenge of multi-tasking (especially given that faster speech rate is usually associated with higher WMC, at least for read speech [36]). If that is the case, the longer phrase lengths that [35]'s speakers produced would be better explained as the result of this artifactual increase in speech rate, rather than compromised WMC, since faster speech is known to inhibit prosodic breaks ([37],[38]). Notably, the individual differences approach we employ below should allow us to avoid task-related strategies like those that might occur when attempting to manipulate WMC directly.

2. WMC & Phrase Length in Read Speech

2.1. Methods

2.1.1. Speech corpus

The speech analyzed here was collected in the context of a previous study reported in [13]. This corpus consisted of speech elicited from 100 native English speakers (42 male / 58 female,

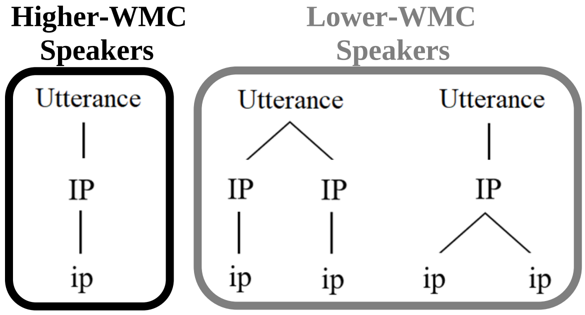

Figure 1: *Schematic representation of hypothesized phrasing preferences in relation to speakers' working memory capacities (WMC). Speakers with lower WMC are predicted to segment utterances into smaller prosodic units at some phrase level, producing shorter phrases on average (either shorter Intonational Phrases (IP), shorter intermediate phrases (ip), or both).*

mostly in their 20s and 30s and metropolitan New York City) who all read the same 156-word passage. The passage (shown in its entirety in the Appendix) was taken from popular prose writing in [39]. Speakers were digitally recorded (44.1k Hz) reading the passage aloud in a sound-attenuated booth using a head-mounted Shure SM10 microphone. All speakers read the passage aloud twice in order to increase familiarity, and especially fluency; only the second production was used for analysis. Recordings were saved as WAV files and set aside for later annotation/analysis of their phrase structure.

Crucially for the present purposes, all speakers in this corpus also completed the Reading Span task, a measure of verbal WMC ([40]), using an E-Prime implementation created by [41]. In brief, on each trial participants were presented with a string of alphabetic letters to hold in memory, and were then required to perform a sentence-comprehension task before recalling the original letter string (in the correct order). Reading spans for each participant were estimated using 'partial-scoring' [42], reflecting the total number of trials with accurate recall (rather than 'absolute' scoring based on a limited number of sets of trials with perfect accuracy), which allows for greater variation across participants to emerge.

2.1.2. Prosodic annotation of phrase structure

In order to determine cross-speaker variation in phrase length, the passage read by each speaker was annotated for its phrase structure using a modified version of the MAE_ToBI transcription conventions ([43]). First, two labelers (working together) identified in each speaker's recording the locations of (a) all disfluencies and (b) all "potential" fluent prosodic boundaries. "Disfluencies" were identified as per the ToBI guidelines and "potential fluent prosodic boundaries" were defined as the locations of any fluent perceived juncture greater than that marking an ordinary word boundary (i.e. anything corresponding to a 2, 3-, 3, 4-, 4, or uncertain 2? or 3?). These first-pass "potential" boundary identifications then served as the input to a more conservative second-pass transcription by two additional ToBI-labelers (working independently), who made final decisions about the break values for the first-pass identifications, assigning them either word-level (0,1,1-, 2, or "?-"), ip-level (3- or 3, or 3?) or IP-level boundaries (4-, 4, or 4?). Agreement levels between the two second-pass labelers, generally consistent with what has been reported elsewhere for the MAE_ToBI conventions ([e.g. 44]), can be found in [13].

The relatively small proportion of disagreements were settled in the direction of the labeler who marked a smaller degree of juncture. That is, for disagreements where one labeler assumed a word-level boundary and the other an ip-level boundary, the word-level boundary was assumed; in cases where one labeler marked an ip-level boundary and the other an IP-level boundary, the ip-level boundary was assumed. After identifying the structure of all IPs for each speaker in this manner, any IP containing a disfluency was excluded from analysis and the length, in syllables, was calculated for all remaining fluent phrases. Across speakers, the length of the average IP was 9.4 (sd=4.1) and the average ip was 6.0 syllables (sd=2.5). Important to point out is that there is overlap between these two phrase types; some IPs (approximately half in this dataset) consist of only a single ip.

2.2. Results

Figure 2 plots each speaker's average ip and IP length from the passage as a function of reading span score (RSPAN; higher RSPAN indicates higher WMC). As can be seen, RSPAN was positively correlated with phrase lengths for both phrase levels, although most strongly in the case of ip, accounting for close to 1/5th of the variance according to a simple R^2.

To test the statistical significance of these correlations in a more rigorous way, mixed-effects linear regression was used to model phrase length, separately for each phrase type. Included in the models as fixed-effects were speakers' RSPAN scores and their mean syllable duration (MSD; calculated across all fluent phrases), the latter being necessary to control for the effect of speech rate on phrasing patterns. (Preliminary inspection of the data revealed that reading spans were inversely correlated with MSD (*r=-.319*), indicating that lower WMC was, in fact, associated with slower speech rate, and thus a confounding factor). Random-effects included those that contributed significantly to model fit as indicated by a log-likelihood ratio test ([45]). These included, in addition to intercepts for "participant" and "sentence in the passage", a by-participant slope for MSD. Continuous variables were centered on their means and adjusted to occur on comparable scales.

Results of the modeling are shown in Table 1 and revealed the following. First, MSD was inversely and significantly related to phrase length for both ips and IPs, indicating that speakers with slower speech rates were, unsurprisingly, associated with shorter prosodic phrases at both phrase levels ($p < .001$). Crucially, however, when this effect was statistically controlled for, RSPAN was a significant predictor of phrase length, such that higher scores were associated with longer ips ($p < .001$), though not longer IPs ($p > .1$).

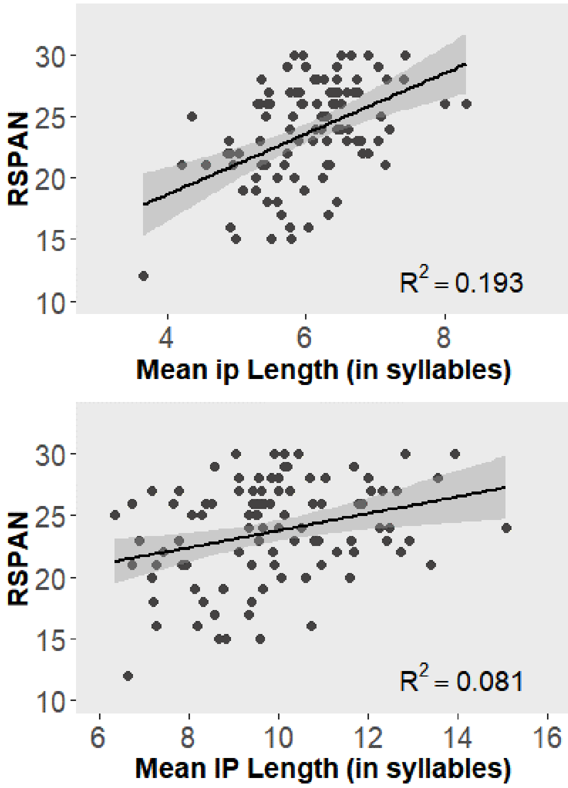

Figure 2: *Mean intermediate phrase (ip) and Intonational Phrase length (IP) as a function of reading span scores (RSPAN) for each speaker. Higher RSPAN indicates higher working memory capacity.*

Table 1: *Results of the mixed-effects linear regression models predicting the lengths of speakers' intermediate phrases (ip) and Intonational Phrases (IP).*

	Estimate	*STE*	*z*	*p*
ip Length (in syllables):				
Mean syllable duration	-.1674	.0214	-7.83	**< .001**
RSPAN	.0284	.0106	2.67	**< .001**
IP Length (in syllables):				
Mean syllable duration	-.3092	.0525	-5.89	**< .001**
RSPAN	.0392	.0275	1.43	> .1

3. Discussion

3.1. Summary of the findings

The question we asked in the present study was whether individual differences in WMC were systematically related to cross-speaker variation in spoken phrase length. The two are hypothesized to be linked by speech production planning; if WMC is one of the cognitive resources required for planning, the planning of speakers with more limited WMC should tend to have correspondingly more limited scope. And, if phonological planning unfolds in prosodically-definable chunks, this limited scope is predicted to have observable consequences for how speakers parse words into prosodic constituents. In fact, we did observe systematic variation of the kind predicted; speakers with lower WMC tended to produce shorter prosodic phrases, and speakers with higher WMC tended towards longer prosodic phrases. We interpret this result as supporting the idea that speakers plan in large, multi-word chunks ([1],[2]), and that the size of these chunks—or at least their length measured in syllables—is sensitive to a speaker's cognitive resources ([20],[27]). In this sense, phonological planning could be said to be prosodic in nature, yet flexible.

Interestingly, however, we found that WMC was the strongest predictor of phrase length at the ip level rather than the IP level. We will now consider this particular finding, which is of particular relevance for modeling prosody's role in speech planning. Before concluding, we also consider the broader implications of the present study for prosody's role in language production more generally.

3.2. Is the intermediate phrase the basic unit of planning?

The fact that we found WMC to primarily predict ip length may suggest a certain primacy for the ip in planning speech, possibly analogous to the primacy claimed for this constituent in sentence processing (see 'Prosodic Visibility' in [46]).

However, as mentioned above, there is very convincing evidence from pause durations and timing adjustments that suggest speakers' planning includes at least basic length information about the entire upcoming IP (e.g. [13],[15]). We therefore think the correlations we found here are better understood in terms of the phrasing choices summarized above in Fig.1. In line with [16], we argue that the relevant aspect of these phrasing options is whether or not they allow for segmented planning to occur. The implication is that both of the two structures on the right of Fig.1 provide this—and thus should be equally good options for speakers under pressure (in this case, speakers with lower WMC resources). And because both of these options increase the number of ips the utterance is parsed into—but only one of them increases the number of IPs—the stronger correlation between WMC and ip length is expected. However, we note that, though it was not significant in the model, a positive correlation between WMC and IP length was also observed, suggesting that the multiple-IP option was likely chosen in some cases by some low-WMC speakers.

An important implication of the present study is therefore that different prosodic structures provide different options for planning, consistent with [16]'s interpretation of her pause duration data. One of our contributions here has been to provide evidence that speakers may actually select their prosodic structures with these planning options in mind (see [47] for insightful discussion of this point).

3.3. Implications for modeling prosodic interfaces

These findings are also relevant to models of prosody's role in language production more generally, as they suggest additional factors outside the language system proper that contribute to prosodic variation. A useful reference for discussion of this issue is shown in Fig.3, based on the diagram by [48], who present it as a starting point for understanding how different factors interact to influence phonetic planning. An important insight of their diagram is its division of factors into those that have a direct effect on phonetic outputs versus those that have a prosodically-mediated one. Non-grammatical factors (e.g. speech rate) are assumed to generally have a direct effect, while grammatical factors (e.g. syntax) generally a prosodically mediated one.

We think the results presented here highlight the need to better understand the contents of the "non-grammatical factors" box, and how they interact with other factors in the language system. A useful way to re-frame the contents of this box may be the way we have framed pressures on speakers—into speaker-external/contextual versus speaker-internal/cognitive types. Interesting to us is that WMC, a clear example of a non-grammatical factor, appears to have an influence on speakers' discrete, categorical phrase structure. We have proposed that the underlying mechanism for this relationship is WMC's importance to speech planning, and so we assume (in line with the prosody-first model) that this implies a prosodically-mediated effect for WMC—and thus the addition of an arrow pointing from "non-grammatical factors" to "prosodic structure". But suffice it to say that the roles and mechanisms of cognitive factors such as WMC, attention, and others are still poorly understood in the context of linguistic modeling. This suggests an important future role for individual differences/psychometric approaches in the investigation of prosody's role in speech production.

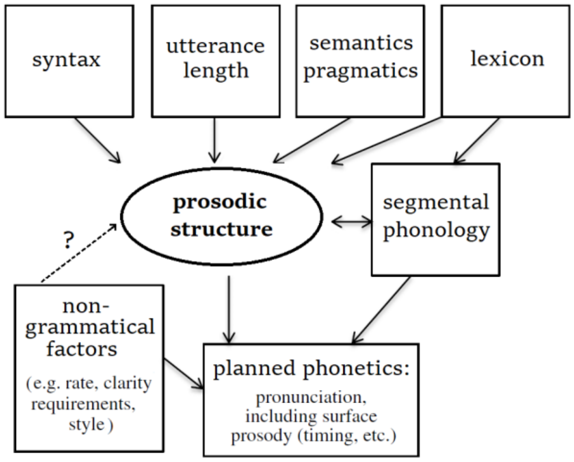

Figure 3: *Modified version of Turk and Shattuck-Hufnagel's (2014) diagram for approaching prosodic interfaces and phonetic planning ([48]). Added here is a dotted line to the "non-grammatical factors" box, representing the possibility of some of these factors (e.g., cognitive processing factors like working memory capacity) to influence phonetic realization via early effects on prosodic structure.*

4. Conclusions

The present study investigated the relationship between phrase length and working memory capacity (WMC) in a large group of American English speakers. Prosodic phrases were defined phonologically and WMC was measured using reading spans. A positive correlation was found. The results are consistent with models of speech planning that allow for flexible scope, but that assume a central, early role for prosodic structure.

5. Appendix: Passage (taken from [39])

Our patch of tangled yard was an exotic foreign country. I had spent so much of my life in dark theatres and dim hotel rooms, where the only thing green was the peeling paint on the walls, that this seemed perfectly natural to me. This was where I had my first bite of mud pie; where I set up a card table and mixed household chemicals, toothpaste, and my mother's face powder, doing what I called "experiments". Now that I was sick I wasn't allowed out of the house, but when I stood on the headboard of my bed I could look through a high window into the backyard and see the concrete wall I used to climb over with my friends. We'd sneak under the cover of the banana trees and light matches I had stolen from my mother's kitchen. I could just reach the big box of wooden matchsticks she kept on the top shelf of the old four-legged gas stove.

6. Acknowledgements

We are particularly grateful to Nadia Zaki for her assistance with the experimental materials, as well as Eve Elliot, Juliana Colon, and Jessica Spensieri for other technical matters. We also thank three anonymous SP2020 reviewers who provided enhancing comments on the original version of this paper.

7. References

[1] P. Keating and S. Shattuck-Hufnagel, "A prosodic view of word form encoding for speech production," *UCLA Working Papers in Phonetics*, vol. 101, pp. 112-156, 2002.

[2] S. Shattuck-Hufnagel, "The role of word structure in segmental serial ordering," *Cognition*, vol. 42, pp. 213-259, 1992.
[3] S. G. Nooteboom, "Limited lookahead in speech production," in F. Bell-Berti and R. J. Lawrence (Eds): *Producing speech: Contemporary issues: For Katherine Harris*, (pp. 3-18). 1995,
[4] D. H. Whalen & J. M. Kinsella-Shaw, "Exploring the relationship of inspiration duration to utterance duration," *Phonetica*, vol. 54, pp. 138-152, 1997.
[5] E. Zvonik, and F. Cummins, "The effect of surrounding phrase lengths on pause duration," in *EUROSPEECH 2003, Geneva, Switzerland, Proceedings,* (pp. 777–780). 2003.
[6] J. Krivokapic, *The planning, production, and perception of prosodic structure*, Ph.D. dissertation, University of Southern California, 2007.
[7] J. Krivokapić, "Prosodic planning: Effects of phrasal length and complexity on pause duration," *J. of Phonetics*, vol. 35, pp. 162-179. 2007.
[8] S. Fuchs, C. Petrone, J. Krivokapić, and P. Hoole, "Acoustic and respiratory evidence for utterance planning in German," *J. of Phonetics*, vol. 41, pp. 29-47. 2013.
[9] G. Kentner, "Length, ordering preference and intonational phrasing: evidence from pauses", *in INTERSPEECH-2007, Antwerp, Belgium, Proceedings*, 2637-2640. 2007.
[10] I. Lehiste, "Interaction between test word duration and length of utterance," *J. of the Acoustical Society of America*, vol. 55, pp. 398–398, 1974.
[11] A. Malécot, R. Johnston, and P. Kizziar, "Syllabic rate and utterance length in French," *Phonetica*, vol. 26, pp. 235-251, 1972.
[12] H. Quené, "Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo," *J. of the Acoustical Society of America*, vol. 123, pp. 1104-1113, 2008.
[13] J. Bishop and B. Kim, "Anticipatory shortening: Articulation rate, phrase length, & lookahead in speech production," in *Proc. of the 9th Conference on Speech Prosody*, pp. 235-239, 2018.
[14] C. Petrone, S. Fuchs, and J. Krivokapić, "Consequences of working memory differences and phrasal length on pause duration and fundamental frequency," in *Proc. of the 9th International Seminar on Speech Production (ISSP)*, pp. 393-400, 2011.
[15] J. Yuan and M. Liberman, "F0 declination in English and Mandarin broadcast news speech", *Speech Communication,* vol. 65, pp. 67-74, 2014.
[16] J. Krivokapic, "Prosodic planning in speech production," in S. Fuchs, M. Weihrich, D. Pape, and P. Perrier, Eds., *Speech planning and dynamics*. Berlin: Peter Lang, pp. 157-190 Speech planning and dynamics, 2012, pp. 157-190,
[17] W. J. Levelt, *Speaking: From intention to articulation*. Cambridge, MA: MIT Press, 1989.
[18] W. J. Levelt, A. Roelofs, and A. Meyer, "A theory of lexical access in speech production," *Brain and Behavioral Sciences*, vol. 22, pp. 1-38, 1999.
[19] W. J. Levelt, "Picture naming and word frequency: Comments on Alario, Costa, Caramazza," *Language and Cognitive Processes*, vol. 17, pp. 299-319, 2002.
[20] V. Wagner, J. Jescheniak, and H. Schriefers, "On the flexibility of grammatical advance planning during sentence production: effects of cognitive load on multiple lexical access," *J. of Experimental Psychology: Learning, Memory, and Cognition*, vol. 36, pp. 423-440, 2010.
[21] M. K. Huffman, "Modulation of speech gestures through prosody or sound change: A commentary," *Laboratory Phonology*, vol. 3, pp. 27-36, 2012.
[22] S. Shattuck-Hufnagel, "Toward an (even) more comprehensive model of speech production planning," *Language, Cognition and Neuroscience*, vol. 34, pp. 1202-1213, 2019.
[23] M. Garrett, "Levels of processing in sentence production," in B. L. Butterworth, Ed., *Language production (Vol. 1: Speech and talk*), London: Academic Press, 1980, pp. 177-220.
[24] S. E. Gathercole and A.D. Baddaley, "Working memory and language," Hove UK: Lawrence Erlbaum Associates, 1993.
[25] R. C. Martin and L. R. Slevc, "Language production and working memory," in M. Goldrick, V. Ferreira, and M. Miozzo, Eds., *The Oxford handbook of language production*, 2014, pp 437-450.
[26] B. Swets, M. Jacovina, and R. Gerrig, "Effects of conversational pressures on speech planning," *Discourse Processes*, vol. 50, pp. 23-51, 2013.
[27] B. Swets, M. Jacovina, and R. Gerrig, "Individual differences in the scope of speech planning: Evidence from eye movements," *Language and Cognition*, vol. 6, 12-44, 2014.
[28] J. Tanner, M. Sonderegger, and M. Wagner, "Production planning and coronal stop deletion in spontaneous speech," *Laboratory Phonology*, vol. 8, article 15, 2017.
[29] L. MacKenzie, "Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation," *Glossa*, vol. 4, article 28, 2019.
[30] M. E. Beckman, and J. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology Yearbook*, vol. 3, pp. 255-309. 1986.
[31] B. Swets, T. Desmet, D. Hambrick, and F. Ferreira, "The role of working memory in syntactic ambiguity resolution: a psychometric approach," *J. of Experiment Psychology: General*, vol. 136, pp. 64-81, 2007.
[32] J. D. Fodor, "Prosodic disambiguation in silent reading," in M. Hirotani, Ed., *Proc. of the North East Linguistics Association (NELS) 32*, pp. 112-132, 2002.
[33] M. Bader, "Prosodic influences on reading syntactically ambiguous sentences," in J. Fodor & F. Ferreira, Eds., *Reanalysis in sentence processing*, Dordrecht: Kluwer, 1998, pp. 1-46.
[34] S.-A. Jun, and J. Bishop, "Prominence in relative clause attachment: evidence from prosodic priming," in L. Frazier & E. Gibson, Eds., *Explicit and implicit prosody in sentence processing*, New York: Springer, 2015, pp. 217-240.
[35] O. Lee and M. Redford, "Verbal and spatial working memory load have similarly minimal effects on speech production," in *Proc. of the 18th International Congress of Phonetic Sciences*, Glasgow, UK: University of Glasgow, paper no. 798, 2015.
[36] M. Daneman. Working memory as a predictor of verbal fluency. Journal of Psycholinguistic Research, 20(6), pp. 445-464. 1991.
[37] C. Fougeron, and S.-A. Jun, "Rate effects on French intonation: phonetic realization and prosodic organization," *J. of Phonetics*, vol. 26, pp. 45-70, 1998.
[38] S.-A., Jun, "The effect of phrase length and speech rate on prosodic phrasing," in M. J. Solé, D. Recasens, & J. Romero, Eds., *Proc. of the 15th International Congress of Phonetic Sciences*, Adelaide, Australia: Causal Productions, pp. 483-486, 2003.
[39] A. Alda, *Never Have Your Dog Stuffed (And other things I've learned)*, New York: Random House, 2009.
[40] M. Daneman, and P. Carpenter, "Individual differences in working memory and reading," *J. of Verbal Learning & Verbal Behavior*, vol. 19, pp. 450-466, 1991.
[41] F. Oswald, S. McAbee, T. Redick, and D. Hambrick, "The development of a short domain-general measure of working memory capacity," *Behavior Research Methods*, vol. 47, pp. 1343-1355, 2015.
[42] A. Conway, M. Kane, M. Bunting, D. Hambrick, O. Wilhelm, and R. Engle, "Working memory span tasks: A methodological review and user's guide," *Psychonomic Bulletin & Review*, vol. 12, pp. 769-786, 2005.
[43] M. E. Beckman and G. A. Elam, "Guidelines for ToBI labelling, version 3.0," The Ohio State Univ. Research Foundation, 1997.
[44] M. Breen, L. C. Dilley, J. Kraemer, and E. Gibson, "Inter-transcriber reliability for two systems of prosodic annotation: ToBI and RaP," *Corpus Linguistics and Linguistic Theory*, vol. 8, no. 2, pp. 277- 312, 2012.
[45] H. Matuschek, R. Kliegl, S. Vasishth, R. H. Baayen, and D. Bates, "Balancing Type I error and power in linear mixed models," *J. of Memory and Language*, vol. 94, pp. 305-315, 2017.
[46] A. Schafer, "Prosodic parsing: The role of prosody in sentence comprehension," Ph.D dissertation, UMASS Amherst, 1997.
[47] F. Ferreira and H. Karimi, "Prosody, performance, and cognitive skill: evidence from individual differences. In L. Frazier and E. Gibson, Eds., Explicit and implicit prosody in sentence processing, New York: Springer, 2015, pp 119-132.
[48] A. Turk, and S. Shattuck-Hufnagel, "Timing in talking: what is it used for, and how is it controlled?," *Philosophical Transactions of the Royal Society B*, pp. 369: 20130395, 2014.