How do Speakers Coordinate Planning and Articulation? Evidence from Gaze-Speech Lags

Chiara Gambi (cgambi@exseed.ed.ac.uk)
Department of Psychology, 7 George Square
Edinburgh, EH8 9JZ UK

Matthew Crocker (crocker@coli.uni-sb.de)
Department of Computational Linguistics and Phonetics, C7.1 University Campus
Saarbrücken, 66123 Germany

Abstract

How do speakers coordinate planning and articulation of more than one word at the same time? Here, we test whether they dynamically estimate how long it takes to (i) plan and (ii) articulate the words they intend to produce as a means of achieving such coordination. German speakers named two pictures without pausing, while their eye-movements were recorded. In line with previous reports, after their gaze left the first picture, speakers took longer to start speaking (i.e., the gaze-speech lag was longer) when the name of the first picture was shorter. But while gaze-speech lags were also longer when the second picture was harder to name, the two effects did not interact. We argue that speakers' flexible planning abilities might be accounted for by reactive, rather than proactive planning mechanisms.

Keywords: planning; estimation; duration; coordination; gaze-speech lag.

Introduction

Speakers plan ahead of articulation (Griffin & Bock, 2000; Levelt, Roelofs, & Meyer, 1999) and usually complete lexico-semantic planning for at least a whole phrase (Martin, Crowther, Knight, Tamborello, & Yang, 2010; Smith & Wheeldon, 1999), phonological planning for a whole word (Meyer, 1996; Smith & Wheeldon, 2004), and articulatory planning for a whole syllable (Meyer, Roelofs, & Levelt, 2003) before beginning to speak. While this allows for rapid and fluent speech production, the incremental nature of planning also raises the question of how speakers manage the timely coordination of planning and articulatory processes.

Mechanisms for Flexible Advance Planning

Several studies have uncovered regularities in speakers' amount of advance planning; for example, suggesting that speech onsets are comparable across short and long words because speakers usually complete articulatory processing for only the first syllable of a word prior to speech onset (Damian, Bowers, Stadthagen-Gonzalez, & Spalek, 2010; see also Meyer, Roelofs, & Levelt, 2003). Importantly, in recent years it has also become clear that the amount of advance planning speakers perform is not fixed, but rather varies with properties of both the utterance and the speaker's recent experience (Konopka, 2012; Van de Velde, Meyer, & Konopka, 2014), or task context (Meyer et al., 2003).

We thus know a great deal about *factors* that can influence the coordination of planning and articulatory processes. On the contrary, we know very little about the *mechanisms* that underlie the timely coordination of these processes. For example, we know that planning style is influenced both by the accessibility of linguistic units, and by the ease-of-apprehension of the referent (Konopka & Meyer, 2014). This suggests that the mechanism involved must be sensitive to the difficulty of the planning process at different stages, but there are at least two ways in which such a mechanism could operate.

One possibility is that a flexible planning system operates reactively. Once speakers encounter some difficulty, a compensatory mechanism is triggered. For instance, if accessing a particular word is difficult (i.e., takes time), attention might be (temporarily) shifted to another process (e.g., retrieving a different word). But in addition, the planning system might, at least in part, allocate resources to different processes in a proactive manner. Such a proactive planning mechanism could be learnt from previous experience with producing language (in general, or within a particular task), and indeed there is evidence that planning style can be primed by previous experience with the same sentence structure (Van de Velde et al., 2014). A proactive planning mechanism would of course be beneficial in maximizing fluency, as it may allow speakers to avoid future difficulties, by anticipating their likely occurrence and taking appropriate steps before they even arise. This idea is reminiscent of some models of motor control (e.g., Wolpert and Flanagan, 2011).

Proactive Planning: Candidate Evidence?

To our knowledge, no language production study has investigated this issue directly. However, in one seminal study, Griffin (2003) suggested that speakers might estimate how long both planning and articulating a word will take, and then combine such estimates to determine how to time one process with respect to the other in order to minimize future disfluencies (i.e., to plan proactively).

To illustrate, imagine a speaker of German preparing to produce *Abschlussballkleider* (dresses for the high-school prom). Let us assume, for the purpose of illustration, that the speaker retrieves *Abschlussball* (prom) and *Kleider* (dresses) separately (Sandra, 1990)[1]. If so, the speaker will need to get the first syllable of *Kleider* ready to be articulated by the time articulation of *Abschlussball* is ending. To do this, the speaker could estimate both how long it will take her to get *Kleider* ready (i.e., retrieval difficulty) and how long it will take her to say *Abschlussball* (i.e., articulation duration). She could then determine that she will probably have enough time to prepare *Kleider* while saying *Abschlussball* (a long word), so she can start speaking right away. But if the first word is short, such as *Sport* in *Sporttitelseite* (sport title page), she might instead have to delay speech onset in order to prepare more of the second word *before* starting to speak. Similarly, she may need to delay speech onset if the second word is particularly difficult to retrieve.

However, the evidence in support of Griffin's proposal is currently somewhat mixed. Griffin (2003) asked speakers to name two pictures one after the other, without pausing, while their eye-movements were recorded. Critically, the name of the picture that was mentioned first (*word$_1$*) could be either short (monosyllabic) or long (plurisyllabic). In this task, speakers usually shift their gaze from the first to the second picture as soon as they have retrieved the phonological representation for *word$_1$* (Griffin, 2001; Meyer & Van der Meulen, 2000). The gaze shift generally occurs before overt articulation of *word$_1$*. Importantly, the interval between this gaze shift and speech onset (i.e., the gaze-speech lag) is longer when *word$_1$* is shorter (a reversed word-length effect). According to Griffin (2003), this shows that speakers estimate *word$_1$* duration: they begin speaking earlier (with respect to the gaze shift, thus leading to a longer gaze-speech lag), when *word$_1$* is shorter in order to have more time to retrieve the second picture's name (*word$_2$*) *before*, rather than *during*, articulation of *word$_1$*.

However, while Meyer, Belke, Häcker, and Mortensen (2007) replicated Griffin's finding[2], they also provided a different explanation. We know that speakers may begin retrieving the articulatory code of the first syllable of a word as soon as they complete phonological processing for this syllable (i.e., without waiting for phonological processing of the whole word to be completed); in turn, as soon as they have retrieved such code, they can begin speaking. But if *word$_1$* is monosyllabic, the moment of the gaze shift (which coincides with completion of phonological processing for the whole word; see above) also happens to coincide with the start of articulatory retrieval. As a result, the gaze-speech lag will last at least the time it takes to perform articulatory retrieval for one syllable. By contrast, for a polysyllabic *word$_1$* the gaze shift occurs only later (once articulatory retrieval of the first syllable is already underway), thus leading to a shorter lag.

This study

If Meyer et al.'s (2007) explanation is correct, then the reversed-length effect on gaze-to-speech lags is not evidence that speakers estimate duration, contrary to Griffin's (2003) suggestion. Moreover, neither study demonstrates that speakers can combine estimates of retrieval difficulty with estimates of duration, because they did not manipulate the difficulty of retrieving *word$_2$*[3]. Here, we provide a test of this hypothesis: If speakers take into account not only *word$_1$* length (monosyllabic vs. polysyllabic words), but also *word$_2$* retrieval difficulty, gaze-speech lags should be affected by both variables. Moreover, the effects of the two variables should interact, reflecting the workings of a proactive planning mechanism underlying the tight coordination of articulation (of *word$_1$*) and planning (of *word$_2$*).

For *word$_2$*, we chose a manipulation that is both known to reliably affect the earliest stages of picture naming, and very easy to identify for participants: Pictures were either visually intact or degraded (see Figure 1). We reasoned this would provide the most favorable test of Griffin's proposal, as speakers were placed in ideal conditions for estimating the difficulty of retrieving *word$_2$*; although degradation does not affect the difficulty of retrieving *word$_2$* directly, it makes accessing the corresponding concept more difficult, which then has a knock-on effect on the time it takes to fully prepare *word$_2$*. To give speakers ample opportunity to adjust to the relevant level of difficulty, and to avoid carryover effects, degradation varied between participants.

As in previous studies, the gaze-speech lag should be longer when *word$_1$* is shorter. In addition, if speakers can estimate retrieval difficulty, it should also be longer when *word$_2$* takes longer to retrieve. Crucially, there should be a significant interaction, with *word$_2$* difficulty having a larger effect when *word$_1$* is shorter. As there is less scope for completing *word$_2$* retrieval during the articulation of *word$_1$* when *word$_1$* is short, speakers should aim to complete most of *word$_2$* retrieval before speech onset; instead, when *word$_1$* is long, the speaker can benefit from extra time after the onset of articulation, and increases in *word$_2$* retrieval difficulty may not affect the gaze-to-speech lag as strongly.

If Meyer et al.'s proposal is correct, however, the gaze-speech lag should only depend on *word$_1$* length, and the reversed-length effect on gaze-speech lags would not be evidence for a proactive planning mechanism. Given the potential theoretical relevance of Griffin's (2003) original interpretation of her findings, testing her claim in full, as we do in this study, would advance our understanding of the

[1] In reality, compound words (especially very frequent ones) might be planned as a single phonological sequence (Jacobs & Dell, 2014). This does not affect the interpretation of our results, as we did not ask our participants to produce compounds, but rather sequences of two unrelated words.

[2] While Griffin (2003) found a reversed word-length effect on speech latencies (i.e., longer latencies when *word$_1$* was shorter) as well as on gaze-speech lags, Meyer et al. (2007) only found this effect on gaze-speech lags.

[3] Although Griffin (2003) varied *word2* frequency and length, between-items differences were very small.

mechanisms underlying flexible planning in language production.

Method

Participants

Thirty-two native speakers of German (24 female, M_{age} = 23.8 yrs, SD = 2.6), with self-reported normal vision and no language impairments, were paid 8 euros/hour to participate in this and another eye-tracking experiment (not reported here). One participant was replaced because of excessive head movements. Sample size was determined on the basis of previous research (Griffin, 2003; Meyer et al., 2007)

Materials

We selected 128 black and white line drawings from the picture naming norms of Bates, et al. (2003). Of these, 64 pictures with high name agreement were used as left pictures. Left pictures were named first, so we refer to the left picture names as $word_1$. For half the items (Long), $word_1$ ranged from 2 to 4 syllables (15 2-syllable words, 11 3-syllable words, and 6 4-syllable words)[4], with a mean length of 2.31 syllables (SD = 0.64). The other 32 pictures had monosyllabic names (Short). Short and long names were yoked in pairs matched for name agreement (Short: .93(.10), Long: .91(.10); t(31)= 1.17, p > .2), log-frequency (Short: 2.62(.46), Long: 2.54(.45); t(31)= 1.51, p > .1) in SUBTLEX-DE (Brysbaert et al., 2011), and initial phoneme.

Sixty-four additional pictures were used as right pictures, and were always named second in the task ($word_2$). Two right pictures were associated with each pair of left pictures. Right pictures had high name agreement (M = .94, SD = .11); $word_2$ had a mean length of 2.14 syllables (SD = .71), a mean frequency of 2.53 (SD = .64), and was semantically and phonologically unrelated to each $word_1$ it was paired with. We created degraded versions of all right pictures by superimposing a mask of ten parallel white lines (about 35pt apart, and about 15pt-thickness; see Figure 1); on average the mask deleted 35% of all non-white pixels (SD = 2.3 %, min = 30%, max = 41%).

Design and Procedure

Length varied within participants and items, whereas Degradation varied within items but between participants. To control for differences due to uninteresting properties of the right pictures, we constructed two lists of items, reversing pairings of left and right pictures (e.g., if in list 1 *Bank* was paired with *Hund*, and *Brücke* with *Krone*, list 2 featured *Bank – Krone* and *Brücke – Hund*); 8 random orders were generated for each list.

[4] Variation in the Long condition was not sufficient to allow treating this variable as a continuous predictor in the analyses. Instead, Length was treated as a categorical predictor (Short vs. Long) throughout.

Participants were first familiarized with picture names. After identifying their dominant eye, they were seated about 60 cm from a 24-inch LCD monitor. A head-mounted Eyelink 2000 recorded data from the dominant eye (pupil-only, sampling frequency: 250 Hz). Participants were asked to avoid head movements and blinking, and named the pictures in left to right order. It was stressed they should avoid pausing between the two words. A high-quality microphone (Philips SBC ME 570) recorded participants' productions for the entire duration of the trial (5.5 seconds); speech onset latencies, and the duration of the pause between names (if present) were then measured offline (in Praat; Boersma & Weenink, 2010).

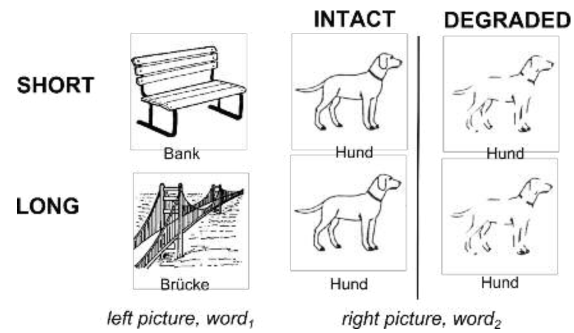

Figure 1: A sample trial illustrating manipulations of $word_1$ length (short vs. long) and $word_2$ retrieval difficulty (intact vs. degraded pictures).

Presentation was controlled using Experiment Builder (Version 1.10.165). Before each trial, a fixation dot was presented where the left picture would subsequently appear. As soon as the participant fixated it, the experimenter triggered presentation of the stimuli (this was also used for drift correction). The left and right pictures were then displayed simultaneously on opposite sides of the screen, 324 pixels (or about 9° of visual angle) apart. All pictures were scaled to a dimension of 290x290 pixels, with surrounding interest areas measuring 405x307 pixels (i.e., 11° of visual angle horizontally, 9° vertically).

The eye-tracker was calibrated twice using a nine-point calibration grid, first after two practice trials, and then halfway through the session. The first trial after the practice session was a warmup trial, and was not analyzed. A session lasted 15-20 minutes.

Results

Only trials in which both pictures were named fluently (i.e., with no repetitions or filled pauses, and with a silent pause no longer than 200ms between the words) and using the expected names were analyzed (intact group: 87.99%; degraded: 83.01%). Following Meyer et al. (2007), we also discarded trials on which the pictures were not fixated in the order of mention (only one trial, degraded group), and trials on which the right picture was not fixated before speech onset (intact: 148 trials, or 16.43%; degraded: 34 trials, or

4.00%)[5], as on such trials the gaze-to-speech lag would have been negative.

For the remaining trials we analyzed speech onset latencies, first-pass gaze to the left picture (the sum of all fixations to the left picture before the shift of gaze to the right picture), and the gaze-speech lag (time between the end of the first-pass gaze to the left picture and speech onset). In all analyses, we fit linear mixed-effects models using the lme4 package (D. Bates, Maechler, & Dai, 2014) in R (R, Version 3.1.3). Fixed effects were contrast coded and centered. Random effects structure was maximal (Barr, Levy, Scheepers, & Tily, 2013). All *p* values are from log-likelihood ratio tests; 95% confidence intervals for model estimates are from the *confint* function (method="Wald"). We report the critical speech-gaze lag analyses first.

Gaze-Speech Lag

As expected, the gaze-speech lag was both *shorter* when $word_1$ was long than when it was short (B=65ms, SE=12, t= 5.56, $\chi2(1)=21.46$, *p*<.001, *CI*=[42,88]) and *longer* for participants naming degraded than intact right pictures (B=-140ms, SE=62, t=-2.24, $\chi2(1)=4.63$, *p*=.031, *CI*=[-262,-17]; see Table 1, top). Crucially, however, there was no interaction between Length and Degradation (B=-14ms, SE=21, t=-.69, $\chi2(1)=0.47$, *p*=.491, *CI*=[-55, 26]).

Speech Onset Latencies

After removing a further 7 (0.40%) outliers (longer than 2500ms), we found speech onset latencies were not affected by Length, whether alone (B=-14ms, SE=27, t=-.50; $\chi2(1)=0.21$, *p*=.645, *CI*=[-67,40]), or in interaction with Degradation (B=5ms, SE=25, t=.19; $\chi2(1)=0.04$, *p*=.846, *CI*=[-44,54]). However, speech onset latencies were longer for participants in the degraded than in the intact group (B=-125ms, SE=60, t=-2.09; $\chi2(1)=4.80$, *p*=.028, *CI*=[-243,-8]; see Table 1, middle).

Table 1: Mean gaze-speech lag, speech onset latency, and first-pass gaze to the left picture, in milliseconds (standard deviation of participants' means in brackets), as a function of $word_1$ Length and Degradation.

Gaze-speech lag		
	Degraded	Intact
Long	437(201)	311(150)
Short	504(210)	362(158)
Speech onset latency		
	Degraded	Intact
Long	1108(169)	985(166)
Short	1102(211)	979(201)
First-pass gaze to the left picture		
	Degraded	Intact
Long	678(77)	699(100)
Short	619(71)	635(100)

[5] Perhaps parafoveal information was sufficient for speakers to identify intact right pictures more often than degraded ones.

First-Pass Gaze to the Left Picture

The time spent looking at the left picture before gaze was shifted to the right was not affected by Degradation, whether alone (B=17ms, SE=27, t=.63; $\chi2(1)=0.35$, *p*=.555, *CI*=[-36,71]; see Table 1, bottom) or in interaction with Length (B=-10ms, SE=24, t=-.43; $\chi2(1)=0.18$, *p*=.668, *CI*=[-58,37]). However, left pictures were fixated for longer if they had long than short names (B=-66ms, SE=25, t=-2.70; $\chi2(1)=7.21$, *p*=.007, *CI*=[-115,-18]), confirming that speakers shifted their gaze as soon as they completed phonological retrieval for $word_1$.

Discussion

We asked speakers to produce fluent two-word utterances and showed that the way they coordinate planning of the second word and articulation of the first word depends on both the length of the first word and the difficulty associated with retrieving the second word. The gaze-speech lag was shorter when participants were preparing to produce a long *word1* and longer when *word2* was harder to retrieve.

However, we found no evidence for an interaction between $word_1$ length and $word_2$ retrieval difficulty. As expected, speakers in both groups took longer to articulate $word_1$ when it was polysyllabic (554ms for the intact group, 539ms for the degraded group) than when it was monosyllabic (401ms for the intact group, 393ms for the degraded group). This difference (about 150ms) is actually larger than the difference in speech onset times between the two groups of speakers (about 125ms). So, speakers in the degraded group *could* have had sufficient extra time *during* the production of a long $word_1$ to compensate for the additional difficulty associated with retrieving the name of a degraded picture. In other words, if these speakers had planned proactively, they could have started speech earlier (with respect to the gaze shift) when $word_1$ was long than when it was short, as only in the latter case delaying speech onset would have benefitted fluency. Had they done so, gaze-speech lags would have been longer for participants in the degraded group than participants in the intact group (as we observed) *but more so* when participants were preparing to produce a short $word_1$, than when they were preparing a long $word_1$.

This is not what we observed. Instead, participants in the degraded group appear to have used a different strategy, delaying speech onset regardless of $word_1$ length. Therefore, a strong version of Griffin's (2003) proposal is ruled out by our findings, as our speakers did not appear to be able to combine estimates of articulation duration with estimates of retrieval difficulty in order to precisely time articulation (of $word_1$) with respect to planning (of $word_2$).

Meyer and colleagues (2007)'s proposal, instead, is compatible with our results. First, it provides an alternative explanation of the reversed word-length effect on the gaze-speech lag, which does not require a proactive planning mechanism. In addition, it may also explain the later speech onsets for speakers in the degraded group, as Meyer et al. (2007) recognized that speakers may not always start

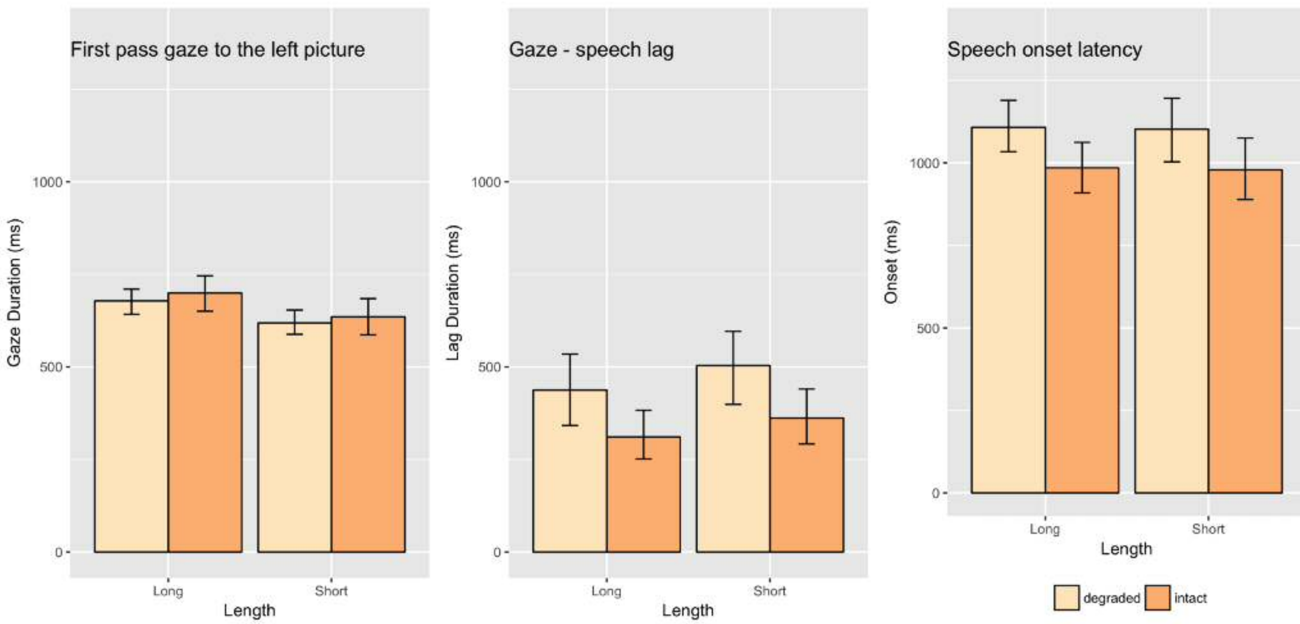

Figure 2: First-pass gaze to the left picture, gaze-speech lag and speech onset latency (means and 95% bootstrap CI), as a function of $word_1$ Length and Degradation.

articulation as soon as the articulatory code of the first syllable of a word has been retrieved.

We suggest that speakers in the degraded group buffered the first syllable of $word_1$ when $word_2$ representations failed to reach some activation threshold sufficiently early, or levels of competition within the production system (see Nozari, Dell, & Schwartz, 2011) remained too high.[6] Importantly, this type of planning mechanism can be considered reactive rather than proactive: It deals with difficulties (with $word_2$ retrieval) as they arise. It need not involve a mechanism that dynamically anticipates the likelihood of future difficulties, deploying different planning strategies depending on this likelihood being higher (i.e., when $word_1$ is short) or lower.

Interestingly, based on our findings, it appears that speech is not planned proactively at the level of whole words. This appears to contrast with what we know about planning at the level of single sounds or syllables (e.g., Hickok, 2012), where there is evidence that speakers build forward models of upcoming speech movements that allow them to anticipate (and quickly correct, if necessary) what they are going to sound like (e.g., Niziolek, Nagarajan, & Houde, 2013).

What might account for such discrepancy? We see at least two possibilities. First, research into forward models for speech has largely focused on speakers' ability to correct a distortion in the spectral properties of the sounds they generate. We are not aware of any studies that investigated whether speakers anticipate and correct for the duration of a sound in a similar way as they do for spectral properties (e.g., pitch).

Second, in order to show the expected behavior under a proactive planning account, our speakers would have had to anticipate not just duration, but also retrieval difficulty. The latter is, unlike duration or pitch, a property of the process of planning itself, rather than an externally perceivable outcome of the planning process. As such, anticipating retrieval difficulty might involve a kind of "second-order" forward model. Speakers might be able to learn such forward models, but perhaps only with extensive training.

In conclusion, the reversed word-length effect cannot be interpreted as evidence that the flexibility of speakers' planning reflects the workings of a proactive mechanism. However, speakers are able to reactively compensate for retrieval difficulty, delaying speech onset when the need arises.

[6] Alternatively, retrieval difficulties with $word_2$ could have interfered with preparation of the articulatory code for the first syllable of $word_1$, and slowed it down. However, note that articulatory retrieval does not appear to impose huge demands on central attention (Roelofs & Piai, 2011). Also, this interference account would have also predicted a smaller effect of degradation on the gaze-speech lag when $word_1$ was long, because with long words the temporal overlap between articulatory encoding for the first syllable of $word_1$ and $word_2$ retrieval should have been shorter.

Acknowledgments

We thank Raphael Morschett for assistance with preparation of the word lists and data collection. This research was supported by research funds assigned to the Chair of Psycholinguistics by the University of Saarland. Chiara Gambi is supported by a Leverhulme Trust Research Project Grant to Hugh Rabagliati and Martin Pickering (RPG-2014-253).

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68(3)*, 255-278.

Bates, D., Maechler, M., & Dai, B. (2014). Lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-7). Retrieved from http://lme4.r-forge.r-project.org/

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., . . . Pléh, C. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10(2)*, 344-380.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 4.6.22) [Computer Software]. Retrieved from http://www.praat.org/

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58(5)*, 412-424. doi:10.1027/1618-3169/a000123

Damian, M. F., Bowers, J. S., Stadthagen-Gonzalez, H., & Spalek, K. (2010). Does word length affect speech onset latencies when producing single words?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 892.

Experiment-Builder. (Version 1.10.165) [Computer Software]. Ottawa, Ontario, Canada: SR Research. Retrieved from www.pstnet.com

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82(1)*, B1-B14.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10(3)*, 603-609.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11(4)*, 274-279.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13(2)*, 135-145.

Jacobs, C. L., & Dell, G. S. (2014). 'hotdog', not 'hot' 'dog': the phonological planning of compound words. *Language, Cognition and Neuroscience*, *29*(4), 512-523.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language, 66(1)*, 143-162.

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1-40.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22(1)*, 1-75.

Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition, 116*(2), 177-192.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language, 35*(4), 477-496.

Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language, 57(2)*, 210-231.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language, 48(1)*, 131-147.

Meyer, A. S., & Van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review, 7(2)*, 314-319.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience, 33(41)*, 16110-16116.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63*(1), 1-33.

R. (Version 3.1.3) [Computer Software]. Vienna, Austria: R Development Core Team. Retrieved from http://www.R-project.org

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology, 2,* 10.3389/fpsyg.2011.00307.

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology, 42(3)*, 529-567.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73*(3), 205-246.

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 675-686.

Van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*(1), 124-144.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current biology, 11*(18), R729-R732.