Journal of Memory and Language 110 (2020) 104066

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The world is not enough to explain lengthening of phonological competitors

Andrés Buxó-Lugo[a,1], Cassandra L. Jacobs[b,*,1], Duane G. Watson[c]

[a] University of Maryland, College Park, United States
[b] University of Wisconsin, Madison, United States
[c] Vanderbilt University, United States

ARTICLE INFO

Keywords:
Prosody
Audience design
Language production
Auditory memory

ABSTRACT

Speakers tend to lengthen the durations of words when a phonologically overlapping word has recently been produced. Although there are multiple accounts of why lengthening occurs, all of these accounts generally assume that competition at some point in the production-comprehension process leads to lengthening. We investigated the contexts that lead to competition and consequent lengthening of target word duration. In three experiments, we manipulated the contexts in which a target word was produced. Speakers produced simple descriptions of animations involving referents that shared initial phonology with another potential referent (e.g., *hat* and *hand*). In Experiments 1 and 2, we manipulated whether the related referent (i.e. *hat*) was named by the speaker themselves, by another person, or was unmentioned. Experiment 2 additionally made speakers aware of competitors in the environment. In both experiments, we found that lengthening does not occur whenever there are two referents in the display that could be confused, even when it is clear that they are confusable. Instead, speakers only lengthened target words when the speaker or another person had named the phonologically related word out loud. Experiment 3 tested whether the task relevance of a phonological competitor influenced naming and found that speakers lengthen words that overlap with previously produced words, even when they were no longer competitors in the display. We propose that an auditory memory account best explains these results and discuss the implications of these findings for other accounts of lengthening.

Introduction

When language-users engage in conversation, they must monitor an abundance of important information about their current context in order to maintain successful communication. Speakers might keep track of what objects or topics have been talked about, predict what objects are likely to be referred to, and may be aware of which words sound similar to one another. These sources of information not only affect *what* a speaker might choose to say, but also *how* they say it. Speakers dynamically change the way they pronounce words depending on the context. These changes are often explained in terms of communicative and mechanistic forces, which are put in motion in response to the demands of the conversational context. For example, speakers tend to shorten or reduce recently produced words (Bard et al., 2000; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Breen, Fedorenko, Wagner, & Gibson, 2010; Buxó-Lugo, Toscano, & Watson, 2018; Fowler & Housum, 1987; Lam & Watson, 2010; Lam, 2014) and lengthen words that partially overlap with recently produced words (O'Seaghdha & Marin, 2000; Sevald & Dell, 1994; Yiu & Watson, 2015). For example, consider (1) below:

(1a) Bond movies are fun because of the gadgets Bond uses.
(1b) Bob's Bond movie collection is quite extensive.

In (1a) the second mention of "Bond" is likely to be reduced because the word "Bond" has been mentioned previously in the discourse. In (1b), the mention of "Bond" is likely to be lengthened because it shares phonological material with a previously mentioned word, i.e. "Bob." Despite the fact that lengthening of phonological competitors is an established effect (Watson, Buxó-Lugo, & Simmons, 2015; Yiu & Watson, 2015), there are still open questions as to what information is relevant to speakers' modulation of speech rate and word durations. More specifically, we ask whether a referent that is saliently present, but not mentioned, should be treated as part of the common ground. Although research on dialogue suggests that physical and linguistic co-presence is often sufficient for referents to be represented in the common ground and to influence speaker choices (e.g. Clark, 1996), we know little about how these factors influence low-level acoustic choices (e.g. word

* Corresponding author at: Brogden Psychology Building, 1202 W Johnson St., Madison, WI 53706, United States.
E-mail address: cassandra.jacobs@wisc.edu (C.L. Jacobs).
[1] Both authors contributed equally to this manuscript.

https://doi.org/10.1016/j.jml.2019.104066
Received 17 August 2018; Received in revised form 4 October 2019; Accepted 7 October 2019

duration).

In this paper, we focus on understanding the contexts in which speakers lengthen word durations. In the case of lengthening, it is generally assumed that competition at some point in the communicative process leads to longer word durations, and lengthening helps mitigate difficulty associated with this competition, for the speaker, listener, or both. Critically, the dominant accounts of word durations (Aylett & Turk, 2004; Bell et al., 2009; Kahn & Arnold, 2015; Sevald & Dell, 1994; Watson et al., 2015; Yiu & Watson, 2015) make predictions as to why lengthening occurs; however, these accounts do not presently make explicit predictions about what factors determine if a referent or referring expression is relevant to speakers' decisions to shorten or lengthen word durations (though see Jacobs, Yiu, Watson, & Dell, 2015). In the next section, we summarize the main accounts of *why* speakers shorten or lengthen the durations of words. We broadly classify these into *communicative* accounts and *mechanistic* accounts. We then discuss cases that these accounts do not explain, as well as the consequences of different potential patterns of results.

Communicative accounts

One explanation of speakers' modulation of word durations is that speakers tailor their utterances to facilitate understanding by their listeners (Brennan & Clark, 1996; Brennan & Hanna, 2009; Brown-Schmidt, 2012; Brown-Schmidt, Yoon, & Ryskin, 2015; Clark & Murphy, 1982; Fussell & Krauss, 1989; Horton & Gerrig, 2002; Yoon & Brown-Schmidt, 2014). In cooperative tasks, speakers readily distinguish between multiple referents when doing so aids listener comprehension. Similarly, speakers may modulate word duration in order to facilitate comprehension. They may lengthen new or unexpected referring expressions that are likely to be difficult to process while shortening predictable, given, or repeated words, which are presumably easier for listeners to process. By some accounts, this tailoring happens because speakers explicitly optimize information transfer to listeners (Aylett & Turk, 2004; Buxó-Lugo et al., 2018; Buz, Tanenhaus, & Jaeger, 2016; Cohen Priva, 2015, 2017; Pate & Goldwater, 2015; Seyfarth, 2014; Tily & Kuperman, 2012). New words tend be high in information load because they introduce referents, events, and ideas that are novel to the conversation and are likely to be communicatively important. Repeated words in a conversation tend to be more predictable, and consequently, convey less information (Aylett & Turk, 2004; Bell et al., 2009). Reducing repeated words and lengthening overlapping words ensures that a relatively constant amount of information is conveyed over time and facilitates processing on the part of the listener.

Supporting evidence for this claim comes from a study by Buz et al. (2016) showing that speakers can modulate word durations for a listener who demonstrates apparent comprehension difficulty. In their study, they tested whether speakers tend to hyperarticulate and lengthen potentially confusable words that overlap phonologically. They showed speakers lists of three words that were either phonologically unrelated (e.g. dim/*cap*/wolf) or contained two phonological competitors (e.g. *gap*/*cap*/wolf). Speakers were shown a highlighted word (e.g. "cap") that they were told to produce for a partner over the internet. Speakers were told that their partner was tasked with selecting that target word. In reality, the partner did not exist, and the participant received experimentally pre-determined artificial feedback about what the virtual partner selected. Occasionally, on trials without phonological competitors, these virtual partners would "select" the wrong word, suggesting that they had difficulty understanding the communication. If these sham interlocutors demonstrated confusion, speakers tended to hyperarticulate (but not necessarily lengthen) words later in the experiment, providing evidence that speakers can flexibly change the acoustic forms of their productions in ways that optimize information load for an audience. The fact that these changes occur in response to a breakdown in communication suggests that audience design plays a role in determining articulation.

An open question for communicative accounts is *when* speakers start to consider information relevant enough to the listener so that they may change their productions. Buz et al. (2016) data suggest that explicit feedback from listeners can guide speakers' choices, but to what extent do listeners anticipate or model listener difficulty without explicit feedback? If speakers are shaping their message for the benefit of the listener, it is plausible that they would consider potential competitors and try to minimize confusion even if the potential competitor has not been overtly mentioned. This has important implications for how speakers form their utterances. Speakers might *broadly* lengthen word durations whenever that word could be confused with another word. If, for example, two interlocutors are discussing a visual display that contains two objects with names that might be confusable (e.g. *hat* and *hand*), speakers might lengthen the word "hand" even if "hat" has never been explicitly introduced to the discourse. Alternatively, speakers might lengthen in a relatively *narrow* set of circumstances, such that only referents that have been explicitly mentioned will affect the upcoming message (i.e. speakers will lengthen "hand" only if "hat" has been previously named). Finally, speakers might need the phonological relationship between two referents to be sufficiently obvious before they consider the two referents competitors. Salience can be easily achieved via naming, which makes the phonological forms of the referents explicit, but could also be achieved by restricting the number of potential referents or competitors that a referring expression might refer to (Buz et al., 2016), or listener feedback (Horton & Keysar, 1996). In sum, research in this tradition has not yet specified what contextual cues drive speakers to acoustically differentiate phonological competitors. Identifying what referents are relevant to audience design will help us better identify the upper bounds of the phonetic and acoustic dimensions of audience design.

Mechanistic accounts

Another possibility is that word duration is modulated by processes related to the mechanics of language production (e.g. Arnold, Kahn, & Pancani, 2012; Bybee & Hopper, 2001). We use the term "mechanistic" in reference to language production mechanisms in a Levelt-like model (Levelt, Roelofs, & Meyer, 1999). In this model, there are distinct levels of representation that proceed in a predominantly serial fashion, with message planning at the top. We can consider communicative theories to mostly live in the message planning layer of the production system, with cascading effects on production, including word choice (e.g. Ferreira, Slevc, & Rogers, 2005; Mädebach, Kurtz, Schriefers, & Jescheniak, 2019). In contrast, mechanistic accounts of duration concern themselves with levels below message planning and are predominantly concerned with how recent and lifelong experience influences production (MacDonald, 2013), including the ease of lexical retrieval and phonological sequencing mechanisms.

In a seminal study, Sevald and Dell (1994) found that speakers who repeated strings of words that initially overlapped in their phonology (e.g. pick-pin-pick-pin-pick-pin) produced the words at a slower speech rate than when they produced strings that overlapped finally (e.g. pick-tick-pick-tick-pick-tick). Sevald and Dell (1994) proposed a model in which the production of one word (*pick*) can slow speakers down during the production of the next (*pin*) because the availability of the sequence of phonemes associated with the first word creates interference during the serial selection of phonemes for the second. Speakers proceed down a familiar path (*pi-*) but then experience conflict that must be overcome when the planned material deviates from the previous production. The production system avoids producing the previous word by requiring slower processing in order to successfully retrieve the target word's phonemes. In contrast, words that overlap finally (e.g. *pick, tick*) do not suffer from interference because the initial material never overlapped.

A similar study by Watson et al. (2015) replicated the results of Sevald and Dell (1994) using compound words with different types of phonological overlap. They found that when speakers produced initially overlapping pairs like *layover-layout*, they produced the non-overlapping morpheme (*-out*) with longer durations than in unrelated pairs (*layout-handover*). Yiu and Watson (2015) translated this paradigm to an event description task, in which participants produced sentences containing two words that phonologically overlapped. Word durations were longer when speakers produced utterances containing two phonologically overlapping referent labels (e.g. "The *hat/cat* shrinks and the *hand* flashes.") relative to unrelated pairs (e.g. *car/hand*). Yiu and Watson (2015) argued that this was evidence that lengthening was driven by production-internal processes. Under their account, speakers experience interference in the phonological sequencing of words that phonologically overlap with a recently produced word, especially for initial overlap, which is critical for phonological planning (Levelt et al., 1999; Levelt, 1989).

Similar to communicative accounts, mechanistic accounts predict that competition will lead to slowed speech rate and/or lengthening, but they have yet to specify when a referent should begin to compete. Whereas lengthening due to phonological cohorts is usually studied in contexts in which a speaker names both competitors (as in Sevald and Dell (1994) and Yiu and Watson (2015)), it is possible that there are other ways for speakers to access the phonological representation of competitors even in the absence of articulation (e.g. Lam & Watson, 2010; Zerkle & Arnold, 2016). For example, the naming of a referent by another speaker might lead to the automatic activation of phonologically overlapping words, which would lead to competition once articulation of a competitor begins (e.g. Sevald & Dell, 1994; Watson et al., 2015). Note that hearing phonologically overlapping words in the cross-modal picture-word interference paradigm tends to *facilitate* speech onsets (e.g. Jescheniak & Schriefers, 2001; Schriefers, Meyer, & Levelt, 1990), but relatively little work has examined how others' productions of phonologically similar words impact word durations. Here also, a better understanding of what makes phonologically related words compete with one another—and therefore lead speakers to lengthen word forms—would provide valuable information in constraining mechanistic theories. Whereas the key question in the context of communicative accounts concerns the information speakers consider when forming a message, the key question for mechanistic accounts is in understanding when words compete with each other.

For both mechanistic and communicative accounts, word lengthening results from *competition*. In the case of communicative accounts, lengthening helps reduce the competition in the mind of the listener between a new referent and a referent that has been recently produced. In mechanistic accounts, lengthening results from and possibly mitigates competition in the speaker's production system that is the result of producing two similar phonological representations. Here, we use the term competition to refer to both types. Rather than adjudicating between mechanistic and communicative accounts, the present study explores the contextual factors that influence whether and how this competition occurs.

The present study

We conducted three behavioral experiments to better identify the constraints on word lengthening in general. Using an event description task paradigm (e.g. Jacobs et al., 2015; Lam & Watson, 2010, 2014; Arnold et al., 2012; Yiu & Watson, 2015; Zerkle & Arnold, 2016), we tasked speakers with producing utterances such as, "The hand flashes." Within each of these utterances, we measured the duration of the noun (e.g. *hand*), which we call the *target*. We manipulated the relationship between the linguistic context (the display) and the to-be-named target. In the experimental conditions for Experiments 1 and 2, a referent whose dominant label overlapped phonologically with the target (e.g. *hat*), which we term the *competitor*, was present in the display. This competitor referent was either produced by the speaker, heard by the speaker, or present in the display but unnamed. In the control condition, the competitor was not present at all, and was replaced with an alternate prime. In Experiment 3, the competitor was either on the same display as the target or on a different display. The mechanistic and communicative theories discussed above currently make no clear predictions about word durations in these three conditions. However, what speakers do in these situations has important implications for these theories. With that in mind, we discuss the experimental outcomes, and how different results might constrain mechanistic and communicative accounts of word lengthening, focusing for the moment on the hypotheses tested in Experiment 1 and 2.

One possibility is that speakers employ lengthening relatively broadly, such that speakers lengthen if there is any potential at all for competition. Under such constraints, a phonological cohort can affect the production of a word if the word was mentioned by the speaker themselves, by someone else, or even simply by being present in the environment. For communicative accounts, such a pattern would suggest that speakers readily consider the potential for confusability when speaking, and so they provide information to avoid confusion even before the confusable information has been explicitly introduced into the discourse. If the speaker has privileged knowledge about the existence of two phonological competitors, this could lead to overly informative lengthening, analogous to naming a "large triangle" without the apparent presence of a small one (Horton & Keysar, 1996). For mechanistic accounts, this would suggest that speakers readily have access to the phonological representations of referents (Lupyan & Swingley, 2011), such that even the presence of a potential competitor might cause competition during naming. There is some evidence that this is true: research using the picture-picture interference task (Jescheniak et al., 2009, 2017; Meyer & Damian, 2007; Morsella & Miozzo, 2002) has found that the presence of a phonological competitor (e.g. a "hat" in a display as a context image along with "hand") can speed naming latencies.

On the other hand, speakers might use narrower constraints on what should count as relevant information when speaking. In this case, speakers should only lengthen words that have been explicitly mentioned by them alone. For communicative accounts, speakers may attempt to disambiguate their utterances, but only once they realize that a referent they have named is confusable for the listener, potentially in response to feedback about a communication failure. This would be similar to the monitoring and adjustment account of audience design (e.g. Horton & Keysar, 1996). For mechanistic accounts, this pattern would suggest that only words that have been explicitly named by the speaker introduce competition into the production process. Some research using the cross-modal picture-word interference paradigm has found *facilitation* of reaction times following the auditory presentation of a phonological competitor (Jescheniak & Schriefers, 2001; Schriefers et al., 1990). However, it has yet to be explicitly tested whether the introduction of a phonological competitor into the discourse by another speaker can lead to interference during articulation and therefore lengthening of word *durations* in the same manner as the introduction of a word into the discourse leads to phonetic reduction on subsequent mentions (e.g. Bard et al., 2000; Kahn & Arnold, 2015).

There is one more possible pattern of results that we could observe. Prior research by Jacobs et al. (2015) demonstrated that overt production of a word was necessary for repetition reduction; simply producing a word in inner speech or mouthed inner speech, which requires speakers to retrieve the phonological form of a word, was not sufficient. Speakers also reduced words when they had previously produced a homophone (e.g. pie-pi) and therefore repeated the same lexeme (Levelt, 1989). In order to account for their results as well as those of Kahn and Arnold (2015) and Bard et al. (2000), who report that reduction also occurs when speakers *hear* a word introduced into the discourse, Jacobs et al. (2015) proposed an *auditory memory* account of repetition reduction. Their proposal states that a referent label must be spoken aloud and that speakers gauge their memory for a phonological sequence to determine how much to reduce. The more familiar the phonological sequence, the faster sequencing can take place, leading to greater reduction on repeated words. The Jacobs et al. proposal puts constraints on repetition reduction by stating that a referent must have been named for it to influence subsequent productions. In some ways, their proposal is intermediate between the broad and narrow proposals that we outlined above, though it accounts only for reduction and not lengthening. It is possible, however, that their proposal could generalize to lengthening as well. If auditory memory is involved in lengthening as it is in reduction, this would provide constraints on what information is relevant for communicative and mechanistic accounts of lengthening.

To better understand the three proposals, we conducted Experiment 1 and Experiment 2, which contrasted these four conditions. In Experiment 1, we tested each of these four conditions, comparing the different types of competitor presence and mention to the absent case. Speakers lengthened words whose competitors had been named aloud but did not lengthen words if a competitor was present but unnamed. In Experiment 2, we wished to rule out the possibility that the lack of lengthening in the *cohort unnamed* condition was an artifact of the task. In this experiment, speakers retrieved the phonological forms of all of the objects on the screen prior to naming the competitors in mouthed inner speech (e.g. Oppenheim & Dell, 2008), with the goal of guaranteeing that speakers were aware of the phonological similarity between the different referent labels they might be producing. Finally, in Experiment 3, we manipulate whether task relevance influences word duration.

Experiment 1

Method

Participants

48 participants were recruited from the University of Illinois paid subject pool. They received $8 for their participation. All participants were native speakers of English who acquired no other language before the age of 5.

Materials

40 critical prime-target pairs were taken from Yiu and Watson (2015) with some additional constraints. To facilitate word segmentation, we excluded items that contained a word-final [ʃ] or [f], with the exception of two items containing the affricate [tʃ] (*wrench* and *beach*). Most words were monosyllabic, though 13 of the targets were disyllabic. The critical stimuli are presented in the Appendix A.

For each display, one image (e.g. hat) was designated the prime, and another (e.g. hand) the target, plus two distractor items. In the no cohort present condition, the trial contained only one critical item (the target) and the prime was replaced by an additional distractor image

Fig. 1. Example critical display from the event description task with phonological competitors “hand” and “hat” present.

(i.e. the “Alternate Prime” column in the Appendix A). In the other three conditions, only two distractors were included on critical trials. On filler trials, all four images were drawn from the set of distractor items.

An additional 40 targets served as filler trials. Each trial contained four potential referents consisting of a prime, a target, and two distractor images. In total, there were 204 unique images (40 critical primes, 40 critical targets, 124 distractors). Filler images were used on every trial type.

In addition to the images, 20 of the trials contained a recording. These recordings were produced by a male native speaker of American English with the same dialect as most participants. The speaker produced simple commands like, “Make the hat shrink.” The speaker spoke in a clear voice and at a natural pace.

Counterbalance

As there are 40 target images and four conditions, we created four fixed lists containing 10 items each, assigning targets randomly to the four lists. We then counterbalanced targets across participants using a Latin square design. Trial orders were randomized for each participant.

Procedure

Participants first went through a training phase to learn the intended labels of each image, a procedure which has been used successfully in other studies (e.g. Trude & Brown-Schmidt, 2012; Yiu & Watson, 2015). Each image was shown along with its intended label at the bottom of the screen. Participants studied the image and its label only once at their own pace before clicking the mouse to continue on to study the next image-label pair. Image-label pairs were presented in a randomized order.

After training, participants were given instructions about the event description task. Trials contained four images, one in each corner. We present an example configuration in Fig. 1. Each trial was broken up into two halves, the prime description and then the target description. Participants were told that two different types of trials would be present in this experiment, which were intermixed. In the three production-only conditions (60/80 of trials), participants would see one image – the prime – shrink (e.g. a hat), wait for a beep to indicate that they could start speaking (500 ms after the end of the shrinking event), and then provide a description of it (e.g. “The hat shrinks”), then click to end the recording. Then, on the second half of the trial (the target phase), another image (e.g. a hand) would flash, followed by a beep (500 ms after the end of the flashing event), and then provide a description of the second event (e.g. “The hand flashes”). After clicking, the next trial would begin with a new set of images. In the fourth condition (20/80 of trials), participants first heard a voice telling them to make one of the items in the display (the prime) shrink (e.g. “Make the hat shrink.”). Participants would click on the prime, which would lead to an animation of the prime image shrinking, and then would see another image (the target) flashing, were cued by a beep 500 ms later, and then produced a description of the event (e.g. “The hand flashes.”). After they finished speaking, participants would click to continue to the next trial with a new display.

Results

Data coding

The second author manually coded the targets on all critical trials for the speech onset times (SOT), durations of the article, durations of the noun, the duration of the verb, and any pauses in Praat (Boersma & Weenink, 2017).[2] Transcription of word boundaries was conducted

[2] We collected all of these measures for completeness in case these measures might be of use to other researchers, but we do not analyze them as they are not relevant to the present research. The data from all experiments are available on

Table 1
Predictions by different theories of whether speakers will lengthen phonological competitors with varying discourse status.

Condition	Example	Broad	Intermediate	Narrow
Competitor said first	Speaker produces "The hat shrinks. The hand flashes."	**Yes**	**Yes**	**Yes**
Competitor heard first	Speaker hears "Make the hat shrink." Then says, "The hand flashes."	**Yes**	**Yes**	No
Competitor unnamed	"Hat" present but unmentioned. Speaker says "The chair shrinks. The hand flashes."	**Yes**	No	No
No competitor present	No hat present. Speaker says "The chair shrinks. The hand flashes."	No	No	No

with both the spectrogram and the waveform information clearly visible. All sound files were anonymized as to condition information by renaming them prior to transcription, which was done to eliminate coding biases during transcription.

Analysis

We considered all targets that had measurable speech onset times (that is, the participant had not started speaking before the start of the recording); the article "the"; nouns that were the intended target; and the verb "flashes." We excluded trials where the participants did not name the targets with their intended labels as determined by the training regime (e.g. *yacht* instead of *ferry*) as well as those with a pause. After selecting for these criteria, 1611 targets remained of the original 1920.

We tested for the effect of having a cohort in the visual display on durations by comparing the three experimental conditions: (1) in which the participant hears someone else name a cohort first, (2) names a cohort first themselves, or (3) where the cohort is present but unnamed, to (4) the baseline condition, where no cohort is present. We modeled these effects using linear mixed effects models in R (*lmerTest* 3.0–1; Kuznetsova, Brockhoff, & Christensen, 2017; *R* 3.4; R Core Team, 2017) with random intercepts and slopes by participants and items for the effect of experimental condition. We chose to use lmerTest as this package provides p values as part of its output using the Satterthwaite (1941) degrees of freedom method. All duration values were centered, scaled and log-transformed for analysis but are presented as raw durations in the figures. The condition predictor was dummy coded with the no cohort present condition as a baseline as this condition is most similar to prior studies on lengthening effects (e.g. Yiu & Watson, 2015).

We found that participants named the cohort targets with significantly longer durations than in the baseline condition *only* when they named the cohort themselves *or* when they heard a recording in which a person named the cohort first. So, if the participants were told "Make the hat shrink", they would say "hand" longer in "The hand flashes." When participants said, "The hat shrinks" and then "The hand flashes", they *also* lengthened "hand", replicating the cohort lengthening effect of Yiu and Watson (2015) and Watson et al. (2015). We summarize these results below in Table 2 and plot them below in Fig. 2.

Discussion

In this experiment, we found that target word durations increased when a competitor was overtly produced either by the speaker or by someone else. This has important implications for mechanistic and communicative accounts of word production, as it suggests that referents must be explicitly mentioned in order for its cohort to be lengthened, therefore providing evidence against the broadest account of lengthening. The lengthening we observe when another speaker produces a phonological competitor could be due to competition during the production process resulting from changes to the production system that hearing a word may cause (e.g. Schriefers et al., 1990) or because

(*footnote continued*)
the Open Science Foundation page (https://dx.doi.org/10.17605/OSF.IO/CK2JG).

Table 2
Estimates of fixed effects from Experiment 1.

	Estimate	Standard Error (SE)	t Value	*p*
Intercept (centered and scaled)	−0.07	0.14	−0.47	n.s.
Cohort unnamed	0.01	0.04	0.25	n.s.
Cohort heard first	0.13	0.05	2.59	$< .05$
Cohort named first	0.18	0.05	3.74	$< .001$

Significance attained at $p < .05$.

speakers become aware of the similarity between referents and adjust their utterances to facilitate comprehension by the listener. The results of Experiment 1 provide evidence against the narrowest account as well. Speakers did not need to articulate the cohort prime label themselves in order for lengthening of the target to occur. Instead, producing or hearing the cohort prime resulted in target lengthening.

Of course, one explanation for the lack of lengthening in the condition in which a phonological competitor is present but unnamed, however, is that speakers were unaware of the phonological similarity between the two confusable referents. We therefore conducted Experiment 2, which added a mouthed inner speech manipulation to ensure that participants retrieved the phonological forms associated with each referent. This manipulation should make the phonological similarity between the two referents obvious. Therefore, if the similarity is relevant for production, we should observe lengthening in all experimental conditions.

Experiment 2

While the results observed in Experiment 1 seem straightforward, it is possible that speakers will lengthen words based on unnamed cohorts if the conditions are right. For example, if a speaker is aware that a listener might confuse two referents, they might choose to lengthen the referent even if the alternative has not been mentioned. Nevertheless, we did not observe this in our data. One potential explanation for this is that participants were not aware that two referent labels overlapped phonologically. Previous work has shown that speakers often recognize the phonological similarity between referent labels after the fact and correct themselves afterward to produce a more informative label (e.g. Ferreira et al., 2005; Horton & Keysar, 1996). Speakers in Experiment 1 therefore might not have had access to the relevant phonological information that would have led to mechanistic-based lengthening. Similarly, speakers might be as informative for listeners as possible, but might not provide helpful cues if they do not realize there is a potential for confusion.

To address the possibility that speakers were unaware of the phonological similarity between the competitor in the display and the to-be-named target, Experiment 2 extends the design of Experiment 1 and introduces a mouthed inner speech naming component to every trial. In this experiment, we explicitly required participants to retrieve the names of all of the referents in the display on each trial before beginning the event description task. We used a mouthed inner speech task (e.g. Jacobs et al., 2015, Experiment 2; Oppenheim & Dell, 2008), which requires participants to retrieve the phonological sequence corresponding to a word without producing the fine motor movements

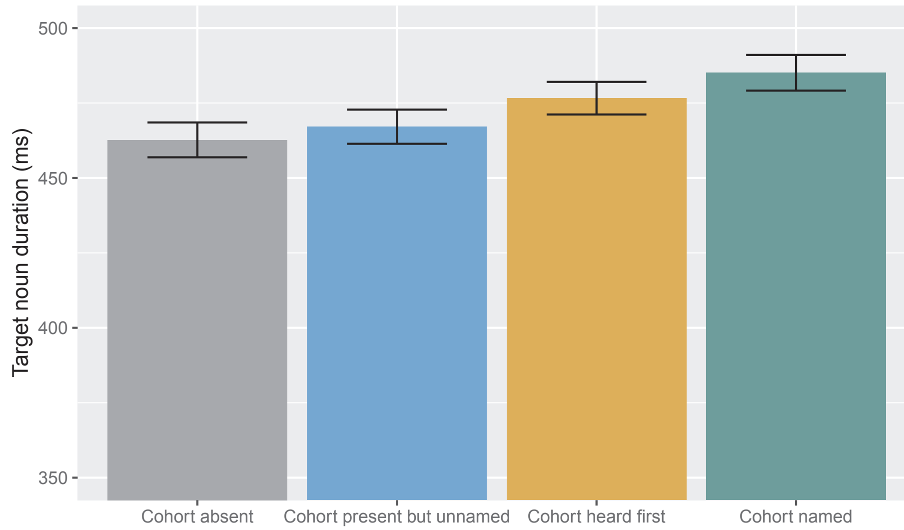

Fig. 2. Means and standard errors for each condition in Experiment 1. Participants significantly lengthened targets when competitors had been named aloud, but not when the competitor was simply present in the display.

associated with overt speech. Retrieving the phonological representations of all of the referents in the display prior to overtly naming a phonological competitor should increase the likelihood that speakers will be able to identify a relationship between the two phonological competitors. If access to the referents' phonological content is enough to cause speakers to distinguish between referents or to create production interference, then we should now observe lengthening in the condition where the cohort is merely present, but never named. On the other hand, if explicit mention is necessary for speakers lengthening due to competitors, we should observe a replication of Experiment 1.

Method

Participants

48 participants were recruited from the University of Illinois course credit subject pool. All participants were native speakers of English who acquired no other language before the age of 5.

Materials

The same materials from Experiment 1 were used for Experiment 2.

Procedure

The procedure for Experiment 2 differed from Experiment 1 in one key way: participants were asked to silently mouth the names of each item on the screen before they witnessed the series of events they were to describe. At the start of each trial, the pictures on the screen were highlighted in clockwise order, starting with the picture on the top-left corner (e.g. hand-corn-hat-brush, as in Fig. 1). Pictures were highlighted with a red square for 2.5 s each. Participants were instructed to look at the highlighted item and mouth the name of the object without saying the name out loud. No label feedback was provided to participants. After naming all items in mouthed inner speech, trials proceeded normally as in Experiment 1 with overt production of both the prime and target. Throughout the course of the experiment, the experimenter verified that participants were mouthing items by periodically observing the participant. Participants who whispered or failed to mouth were reminded of the instructions.

Results

Data coding

Recordings were manually segmented into words by the first author and a research assistant using Praat (Boersma & Weenink, 2017) with both the waveform and spectrogram fully visible. Each recording was segmented into up to five regions: the speech onset, the determiner, the target noun, the verb, and a following pause.

As the initial coding was unintentionally not blind to condition, the second author re-coded a portion of the participants (13 of the final 38) using the same guidelines but for de-identified files. We compared mean durations for the items that met our inclusion criteria below (n = 473). There were no significant differences in overall durations between the coders (intercept = 0.002, $t(471) = 0.30$, p = n.s.), and the correlations between the second author's de-identified durations and the original durations were very high ($\rho = 0.94$, $t(471) = 60.68$, $p < .001$). Given the high agreement between the anonymized and original annotations, we choose to analyze only the original annotations for consistency.

Analysis

We again tested for the effect of having a phonological competitor in the visual display on durations by comparing the three experimental conditions to the baseline condition (the displays in which no phonological competitors are present). Following the same criteria as in Experiment 1, we excluded trials where participants named the target with the unintended label (132). We also excluded participants who did not reliably produce mouthed inner speech (defined as any individual who needed to be told to mouth the referents prior to the start of the trial twice or more; n = 6) and recording issues (n = 4) due to background noise. After these exclusions, 38 participants and 1300 trials remained. As in Experiment 1, we model the effect of condition on word durations using linear mixed effects models in R (*lmer* version 1–1.13; *R* version 3.4) with participant- and item-level random intercepts and slopes for the effect of experimental condition. Condition was dummy coded with a control baseline.

Our results closely resemble those from Experiment 1. Participants produced the competitor target labels with significantly longer durations than in the baseline condition only when (1) they had themselves named a word in the same phonological cohort *or* (2) when they first heard a recording in which another person named the cohort. There was no significant difference between the durations of targets that had a phonological competitor present in the display and the baseline condition. In other words, even though participants had to access cohorts' phonological form in order to mouth their names at the beginning of the trial, this was not enough for them to reliably lengthen the target. We note that average durations in Experiment 2 are shorter than those in Experiment 1 by approximately 50 ms. We can only speculate that

Table 3
Estimates of fixed effects from Experiment 2.

	Estimate	Standard Error (SE)	t value	p
Intercept (centered and scaled)	−0.08	0.14	−0.59	n.s.
Cohort unnamed	0.04	0.05	0.68	n.s.
Cohort heard first	0.19	0.06	3.31	$< .01$
Cohort named first	0.14	0.06	2.54	$< .05$

Significance attained at $p < .05$.

repeating the labels of all the referents after producing them in mouthed inner speech facilitated production mechanisms, resulting in reduced target word durations, as predicted by some theories linking production facilitation and word durations (e.g. Kahn & Arnold, 2012). We summarize these results below in Table 3 and plot them below in Fig. 3.

Discussion

In Experiment 1 we found that speakers lengthened words that they had either explicitly introduced into the discourse or had heard another speaker produce. Speakers did not lengthen potential competitors if a phonologically similar word had not already been mentioned. One possibility for this pattern is that speakers had not retrieved the labels of the other items and did not recognize that they were potential competitors. In Experiment 2, we addressed this concern by instructing participants to mouth the name of each referent before the events took place. This manipulation should have increased the likelihood that participants were aware of the phonological properties of all items present for each trial. Instead, the results from Experiment 2 replicated those from Experiment 1: speakers lengthened words when a competitor has been explicitly mentioned, but not when a potential competitor was merely present, even when speakers had accessed the competitor's phonological representation. Speakers are selective in when they choose to lengthen words (i.e. insensitive to the presence of an unnamed phonological competitor), but not so selective that they only lengthen following their own productions.

The results of Experiments 1 and 2 therefore provide strong evidence for the intermediate account we put forth in Table 1. The intermediate hypothesis stated that speakers do not simply lengthen whenever a phonological competitor is in the environment (the broad account), and do not only lengthen whenever their own production system must contend with producing an utterance that is similar to a recently produced one (the narrow account). Instead, these results show that speakers lengthen so long as an auditory image – produced or heard by the speaker – is available to them. The results also suggest that the auditory memory hypothesis of Jacobs et al. (2015) extends beyond repetition reduction to include lengthening of phonological competitors.

Experiment 3

According to the auditory memory hypothesis, repetition reduction is due to the activation of a phonological form in memory. However, a concern with the experimental design of Experiments 1 and 2 is that previous mention of the prime is potentially confounded with task relevance. Rather than auditory memory of the prime driving the target lengthening effects in Experiment 1 and 2, it is possible that any prime that is relevant to the task at hand is more likely to be activated in the production system and will lead to lengthening of the target. Words that have been previously mentioned may simply be more task relevant than a referent that has not been overtly mentioned.

To rule out the possibility that the effects in Experiments 1 and 2 were driven by task relevance, we altered the paradigm so that primes could be overtly mentioned, but were not relevant to the task on the current trial. We presented critical (phonologically overlapping) primes on either the same displays (*same-display* condition) as their targets or on the previous display (*different-display* condition). As in the previous experiments, we also manipulated phonological overlap between the prime and target.

There are two general patterns of lengthening that we might expect to see in Experiment 3. First, in line with the auditory memory account of Experiments 1 and 2, any availability of a phonological competitor will lead to the lengthening of target word durations, independent of whether the competitor is mentioned in the same display or a different display. This would provide support for the auditory memory hypothesis. Another possibility is that speakers will lengthen words only when it is relevant to do so for the task; that is, only in the same-display (task-relevant) overlap condition, with no lengthening when the prime and target are presented on different displays. This outcome would be consistent with task relevance driving the effects.

Method

Participants

We recruited 55 participants from the Vanderbilt course credit and community subject pools. 37 individuals received one hour of course credit; 11 paid participants received $5 USD for half an hour of their participation. We excluded seven participants: three due to a microphone error, as well as four participants who consistently failed to

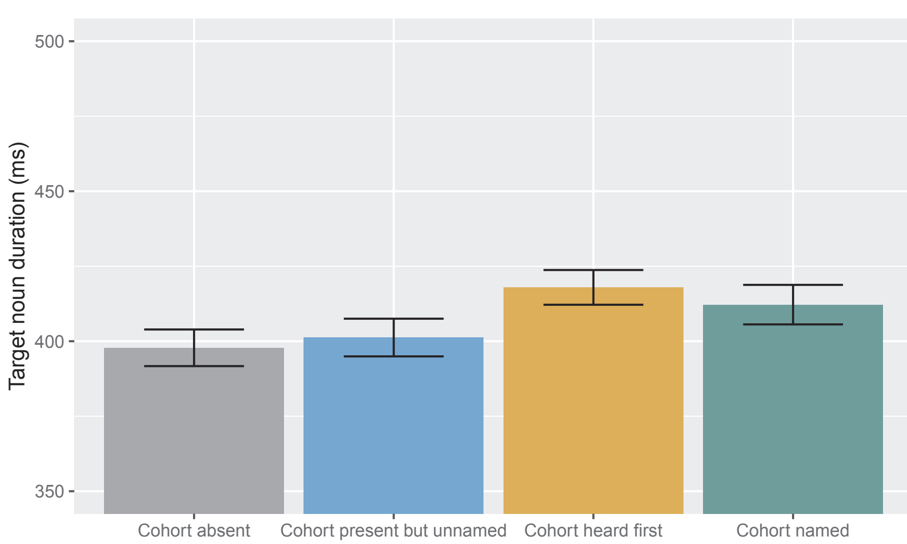

Fig. 3. Means and standard errors for each condition in Experiment 2. Participants significantly lengthened targets when competitors had been named aloud, but not when the competitor was simply present in the display, even though they had retrieved the phonology associated with both the competitor and the target after mouthed inner speech.

follow the instructions throughout the task, for a total of 48 participants for analysis.

Design

The design follows a 2 × 2 factorial design, with the two factors being same/different displays and phonological overlap/phonologically unrelated prime-target pairs. As in Experiments 1 and 2, participants viewed a given target in only one of the four conditions. Participants saw one quarter of the targets per each of the four conditions, which we counterbalanced using a Latin square design. Participants were assigned to a particular counterbalanced list (1–4) based on their participant number, which determined the allocation of a prime-target pair to its condition. Trial order was randomized for each participant.

Materials

Prime-target pairs were the same as those from Experiments 1 and 2. As in Experiments 1 and 2, when a critical trial was a non-overlap trial, the phonologically overlapping prime was replaced with a phonologically unrelated "alternate prime." In Experiments 1 and 2, each trial contained two distractor images. However, the same/different display manipulation of Experiment 3 required creating two displays containing up to three distractor images for the different-display condition. This is because the different-display condition splits a single trial into two consecutive one-event "trials" in which the prime is the first event and the target is the second event. Filler trials and critical trials were both subject to the same/different-display manipulation.

When primes and targets appeared on different displays, two changes were made. First, it was necessary to select an additional distractor image for the prime displays, which replaced what would later become the target image. Second, we selected 3 additional distractor images for the target displays in the different display condition. All non-target images were random assortments of unique distractor images. All conditions were counterbalanced across participants using a Latin square as described above. We provide these materials below in Table 2 of the Appendix A.

Procedure

The procedure closely followed Experiments 1 and 2. As before, participants were given a brief introduction to the task and then entered a self-paced study phase in which each image (including filler images) was presented along with its intended label in a random order. This procedure is critical for preventing data loss in production experiments. After this, participants read additional instructions telling them about how recording in the task would proceed.

As before, each trial is split up into two phases, in which one image shrinks and another image flashes. Participants observe the animation for 1 s, are alerted to begin speaking following a half-second millisecond 880 Hz beep, and describe the images using simple sentence frames (i.e. "The hat shrinks" and "The hand flashes"). Participants clicked the mouse to end recording and begin the next animation. Trial order was randomized by participant.

On same-display trials, both control and overlap primes and targets were described in the normal manner following the procedure in Experiment 1 and 2. However, on different-display trials, each trial was broken up into two "trials" in which participants saw two consecutive displays of 4 images and only provided a single description per scene. On the first of these "trials", participants produced descriptions of the "shrinking" event. After clicking the mouse to continue, a new display appeared containing four unseen images, including the to-be-named target. The display remained static on the screen for 1 s, after which the animation began. Again, once the flashing animation for the target image in the new display ended, a beep played, participants began speaking, and then ended the recording with a click to move on to the next trial.

Coding

In order to speed transcription, audio files were force aligned using the Python program Gentle (https://lowerquality.com/gentle/), which is a speech recognizer trained on conversational speech. These alignments were then translated into TextGrids using the program praatio (Mahrt, 2016). The two first authors completed transcription for 20 and 28 participants, respectively. Transcriptions were corrected from the forced aligned TextGrids using both spectrograms and waveforms in Praat (Boersma & Weenink, 2017). During the correction process, the annotators identified and adjusted word onsets, offsets, and marked for pauses, incorrect names, and disfluencies. Annotators were blind to condition as the four condition labels were assigned arbitrary numeric values.

Data exclusion

We excluded all trials on which participants produced anything other than the studied label (e.g. "object" for "racket"), including speech errors or the prime in the phonological overlap conditions was not the intended label (e.g. "castle" instead of "fort"), which would make the phonological overlap manipulation irrelevant. After excluding these recordings, a total of 1670 target files remained for analysis.

Results

We constructed a linear mixed effects model on centered and scaled log target word duration as a function of each of two levels of the two factors (overlap/control and same-display/different-display) and the interaction between these two factors. All factors were sum (−1/1) coded for the purposes of the main analysis. The model included maximal random intercepts and slopes by both participants and items.

The model revealed a significant main effect of prime-target overlap, but no significant effect of display type and no significant interaction between these two factors. We present the results of the statistical model below in Table 4 and plot the mean durations by condition below in Fig. 4.

While the main analysis suggests that lengthening occurs to roughly equal extents in both overlap conditions, we conducted additional statistical tests using linear mixed effects models following the same procedure above but conducted post-hoc comparisons within the same-display and different-display conditions of the experiment, testing for the effect of overlap on scaled log target durations. Both analyses revealed that overlap leads to lengthening, i.e. for both scene types. We present the results of the two models calculated over these subsets below in Tables 5 and 6.

Discussion

The goal of Experiment 3 was to determine whether phonological overlap only impacts whole-word phonetic duration when it is relevant to the specific task. In Experiments 1 and 2, we found that speakers only lengthened the names of referents when phonological competitors were named aloud, even when speakers were aware of unnamed competitors. In Experiment 3, speakers named all primes overtly, but the prime was either present in the discourse within the same display as the eventual target ("hand") or was part of a different display (the previous trial) and

Table 4
Estimates of fixed effects from Experiment 3.

	Estimate	Standard Error (SE)	t Value	*p*
Intercept (centered and scaled)	0.01	0.13	0.05	n.s.
Overlap	0.07	0.02	4.14	$< .001$
Display type	−0.01	0.02	−0.70	n.s.
Overlap × Display type	−0.01	0.01	−0.55	n.s.

Significance attained at $p < .05$.

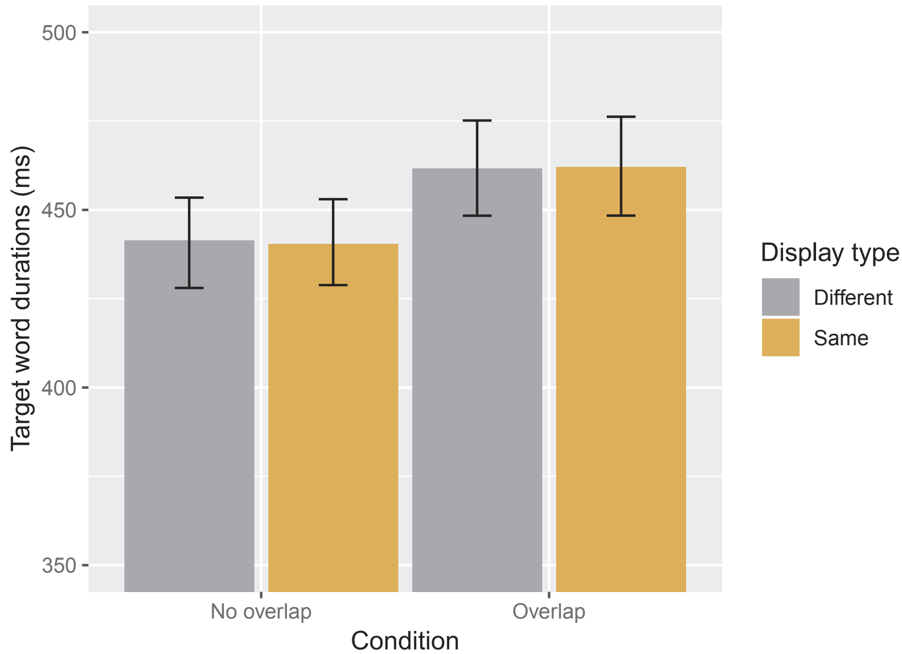

Fig. 4. Mean durations and standard errors by overlap type and display type.

Table 5
Estimates of overlap fixed effect from Experiment 3, same-display condition.

	Estimate	Standard Error (SE)	t Value	*p*
Intercept (centered and scaled)	0.001	0.13	0.01	n.s.
Overlap	0.06	0.02	3.20	< .01

Significance attained at p < .05.

Table 6
Estimates of overlap fixed effect from Experiment 3, different-display condition.

	Estimate	Standard Error (SE)	t Value	*p*
Intercept (centered and scaled)	0.01	0.13	0.10	n.s.
Overlap	0.07	0.02	3.54	< .01

Significance attained at p < .05.

was no longer task relevant. If the relevance of the competitor modulates phonological competition, we reasoned that lengthening would only occur in the same-display overlap condition. However, we found that lengthening of phonologically overlapping target words like "hand" occurred to roughly equal extents independent of whether "hat" was named in the previous display or on the same display as the target. These results strongly suggest that words that have been recently produced, even if they are no longer part of the same visual discourse, are still relevant to the production process. Taken together with the results of Experiments 1 and 2, there is now stronger evidence that these lengthening effects are driven by the availability of an auditory image in memory, and not by the availability or presence of the particular items presented in a display.

General discussion

Speakers are sensitive to the information in their environment, and this sensitivity has specific consequences for productions, e.g. the forms of referring expressions, speech rate, and even fine-grained phonetic forms. Both communicative and mechanistic theories of word production assume that speakers lengthen words to minimize the competition between two or more names, either for the listener, who must disambiguate referents quickly and accurately, or for themselves to minimize the likelihood that they will say the wrong word. In this work we asked what factors influence speakers' word lengthening decisions, as word duration has been shown to be a reliable cue to discourse status (Brown, 1983; Fisher & Tokura, 1995; Haviland & Clark, 1974), production difficulty (Watson et al., 2015; Yiu & Watson, 2015), and may be integral to successful communication (Aylett & Turk, 2004; Buz & Jaeger, 2016).

In the first two experiments, we demonstrated that speakers do not change their pronunciation just because there are two referents with similar labels in a display. Speakers lengthened words that phonologically overlapped with previous words only when the speaker themselves or another person had produced a similar-sounding label for a referent. It appears that speakers do not lengthen words indiscriminately, but instead, are sensitive to what they and others have said. Even when speakers know that there are two similar-sounding referents in a display (Experiment 2), they do not significantly change the durations of the words they produce. This suggests that when speakers must produce a new word, their prosodic decisions are influenced by the referents that are already in the discourse. Simply having a similar-sounding referent in the environment is not enough to influence speakers' productions. It is also not necessary for that referent to be task

Table A1
Stimuli and counterbalance, Experiments 1 and 2.

Set	Target	Prime	Alternate Prime	Filler 1	Filler 2
1	Pillow	Pilgrim	Drum	Wrench	Seal
1	Clock	Claw	Whistle	Corn	Pot
1	Plate	Plane	Kangaroo	Mouse	Sandwich
1	Ferry	Ferret	Church	Harp	Frog
1	Chain	Chair	Nose	Lettuce	Gun
1	Nest	Neck	Glove	Screw	Lightbulb
1	Rattle	Radish	Helicopter	Zebra	Hammer
1	Candle	Candy	Television	Deer	Clown
1	Beach	Beak	Watermelon	Ring	Pear
1	Cake	Cape	Piano	Bird	Doll
1	Ball	Helicopter		Tiger	Glove
1	Train	Kangaroo		Nose	Bed
1	Chicken	Whistle		Zebra	Lightbulb
1	Turtle	Ruler		Frog	Pear
1	Drum	Seal		Ring	Fence
1	Hammer	Spider		Cow	Pencil
1	Thread	Skunk		Flag	Corn
1	Key	Lips		Bird	Motorcycle
1	Watch	Swing		Nail	Gun
1	Fox	Fly		Snake	Button
2	Crater	Cradle	Key	Vest	Ruler
2	Brick	Bridge	Swan	Chain	Strawberry
2	Rose	Rope	Lightbulb	Football	Deer
2	Dime	Dice	Wrench	Truck	Cow
2	Shade	Shark	Chicken	Lettuce	Frog
2	Goat	Goal	Piano	Doll	Spider
2	Bone	Bow	Cat	Ring	Leaf
2	Carrot	Carriage	Bird	Clown	Flag
2	Horn	Horse	Giraffe	Drum	Seal
2	Bat	Bag	Shoe	Fox	Church
2	Grasshopper	Pipe		Spider	Strawberry
2	Cow	Bread		Glove	Leaf
2	Pot	Clown		Tomato	Tiger
2	Harp	Pencil		Corn	Book
2	Nail	Lettuce		Swing	Thread
2	Piano	Seal		Cat	Motorcycle
2	Knife	Church		Lips	Fly
2	Flag	Train		Sandwich	Screw
2	Wrench	Pear		Barn	Tie
2	Fence	Watermelon		Bed	Turtle
3	Saddle	Salmon	Tie	Rooster	Knife
3	Racket	Rabbit	Whistle	Clown	Glove
3	Hook	Hood	Kangaroo	Bird	Pipe
3	Mouse	Mouth	Fly	Motorcycle	Key
3	Mold	Mole	Plug	Skunk	Brush
3	Shower	Shovel	Leaf	Piano	Ball
3	Bell	Belt	Strawberry	Harp	Book
3	Money	Monkey	Drum	Tomato	Button
3	Drill	Drum	Watermelon	Lips	Bed
3	Mustard	Mustache	Deer	Spider	Tiger
3	Star	Deer		Strawberry	Screw
3	Fox	Church		Chicken	Piano
3	Glove	Pot		Knife	Doll
3	Button	Whistle		Swan	Skunk
3	Watermelon	Hammer		Harp	Clown
3	Tie	Vest		Thread	Door
3	Swing	Key		Pencil	Leaf
3	Gun	Fly		Bread	House
3	Shoe	Mouse		Book	Tiger
3	Grasshopper	Lightbulb		Rooster	Car
4	Scale	Scarf	Rooster	Book	Star
4	Fork	Fort	Grasshopper	Truck	Clown
4	Locket	Locker	Chicken	Ring	Key
4	Cable	Caveman	Pipe	Hammer	Fly
4	Hand	Hat	Helicopter	Corn	Brush
4	Wine	Wire	Frog	Chain	Motorcycle
4	Toad	Toast	Peach	Football	Nose
4	Clown	Cloud	Drum	Swing	Gun
4	Pin	Pig	Church	Turtle	Bird
4	Basket	Badger	Shoe	Tomato	Tie
4	Skunk	Ball		Pear	Train
4	Hammer	House		Deer	Clown
4	Television	Piano		Brush	Watermelon
4	Rooster	Watch		Vest	Peach

(*continued on next page*)

Table A1 (*continued*)

Set	Target	Prime	Alternate Prime	Filler 1	Filler 2
4	Tomato	Barn		Nail	Ruler
4	Thread	Glove		Snake	Star
4	Door	Fox		Car	Book
4	Tiger	Cat		Truck	Wrench
4	Football	Turtle		Bread	Lettuce
4	Flag	Key		Swing	Pot

relevant for it to have an effect. Experiment 3 demonstrated that phonologically similar primes words that are no longer in the same visual display as a target cause the same amount of lengthening on target words as when they are in the same display. Altogether, the results suggest that speakers are in fact highly sensitive to what has been said in the past, even if it is not directly relevant to the task, and even when there are potentially confusable referents in the environment (Ferreira et al., 2005).

In summary, our results suggest that there is something special about explicitly mentioning a competitor that mere presence does not achieve, even when the phonological content has been retrieved and should be available. Mouthing in our experiment does not introduce a referent into the common ground, unlike overt production. As such, the potentially similar-sounding referent is only a phonological competitor from the speaker's point of view once the other name has been explicitly negotiated with another person (e.g. Brennan & Clark, 1996).

This is important when considering the broader communicative context of referring expression production. Outside of experiments where speakers must name items with high name agreement, speakers are typically free to conceptualize events and choose names for referents as they please (Quine, 1968). Thus, from the speaker's perspective, two referents are not necessarily competitors before one or both labels have been selected, even though their phonological forms may be active prior to speaking (Mädebach et al., 2019).

What are the implications for the communicative and mechanistic theories of word duration that we discussed in the introduction? These data suggest that communicative theories must include a role for previous mention when modeling speakers' choice to lengthen a word. These theories must also explain why lengthening might be optimal for listeners only when words are previously mentioned. This may be relatively straightforward in contexts similar to Experiments 1 and 2. Words that have been previously mentioned in the discourse are highly accessible to listeners, and consequently, lengthening a similar sounding target word may facilitate comprehension for listeners, by smoothing the information signal at a point of ambiguity (e.g. Aylett & Turk, 2004). The data from Experiment 3 is more difficult to account for within communicative frameworks. In this experiment, the prime word was, at times, not present in the display, and consequently, the referent of the target word was always unambiguous. It is not clear how lengthening the target word in this context facilitates processing by the listener. Doing so may even violate the Gricean maxim of relevance (Grice, 1975). One possibility is that listeners do not model listener representations, but instead, use heuristics (Jacobs et al., 2015). We return to this possibility below.

On the other hand, communicative accounts are not the only possible means of explaining why speakers slow down after producing phonologically overlapping words. Many models of language production (e.g. Sevald & Dell, 1994; Watson et al., 2015) that focus on the dynamics involved in phonological sequencing are not explicitly sensitive to contextual information. It is unclear exactly how the *overhearing* of a word being produced would affect the phonological sequencing process, and in particular why slowdowns arise after hearing phonological competitors. It is especially puzzling that speakers should slow down when overhearing phonologically similar words when cross-modal tasks have shown that hearing phonologically related words can shorten the amount of time needed to initiate speech (Jescheniak & Schriefers, 2001; Jescheniak et al., 2009; Meyer & Damian, 2007). Some models do account for comprehension-production interactions (e.g. Levelt et al., 1999). Still others account for speech rate as a function of phonological factors and recent experience (e.g. Sevald & Dell, 1994), which conflates speech onset latencies and durations. Others are models of speech rate that keep recently produced words in "memory", but not recently *comprehended* ones (Watson et al., 2015). Consequently, no computational account of speech rate thus far properly specifies the ties that bind two phonologically similar words together independent of whether a phonological sequence or word was heard or spoken. Our results show that all mechanistic accounts of the production process must take into account the importance of comprehended input for phonological sequencing.

Both communicative and mechanistic accounts can potentially be amended with the addition of heuristics. For example, Jacobs et al. (2015) proposed that memory for a label can drive phonetic reduction and achieve what the authors termed "audience design by proxy." According to this account, speakers use their own facilitation or interference as a forward model to accomplish audience design at the phonological level without explicitly modeling the internal state of their listener. This theory is similar in spirit to proposals in which speakers egocentrically model common ground by assuming that listeners' representations of the discourse model match their own (Epley, Keysar, Van Boven, & Gilovich, 2004). The sequencing mechanisms outlined in Jacobs et al. (2015) also provide a means for talker-independent phonological experience to impact a speaker's production system. Whenever *any* auditory memory is available to a speaker, the phonological sequencing process will be facilitated up until the point at which the new phonological sequence starts to diverge from the previous one.

Altogether, our results impose important but tractable constraints for mechanistic and communicative theories of prosodic production. First, speakers do not appear to lengthen whenever possible, but in fact fairly selectively, though not so narrowly as to ignore the discourse status of referents or others' productions. Both verbal and computational theories of audience design (e.g. Aylett & Turk, 2004) should therefore be sensitive to the linguistic forms that have been used to refer to different entities in a discourse when predicting whether speakers will lengthen word durations. Secondly, speakers' prosodic decisions can be influenced simply by hearing a similar-sounding word. While there has been a vast literature within the field of language production dedicated to the mechanics underlying reaction times and speech rate (Levelt et al., 1999; Schriefers et al., 1990; Sevald & Dell, 1994; Watson et al., 2015), relatively little work has been dedicated to explaining word durations, and most of the work tends to be insensitive to the importance of memory and availability of phonological competitors.

It is clear that by applying straightforward modifications or constraints that emphasize the role of memory on both mechanistic and communicative theories, the field will come closer to understanding how and why speakers change how they pronounce words more generally. Future work should explore how speakers are able to access their memory for the discourse in making their prosodic decisions to lengthen or reduce, the form and nature of that memory, as well as how mechanistic and communicative goals might differentially incorporate

Table A2
Stimuli and counterbalance, Experiment 3.

Set	Target	Prime	Alternate Prime	Copresent Filler 1	Copresent Filler 2	Copresent Filler 3	Separate Filler 1	Separate Filler 2	Separate Filler 3
1	Pillow	Pilgrim	Drum	Wrench	Seal	Fence	Lips	Swan	Tomato
1	Clock	Claw	Whistle	Corn	Pot	Gun	Thread	Vest	Wrench
1	Plate	Plane	Kangaroo	Mouse	Sandwich	Strawberry	Strawberry	Ring	Zebra
1	Ferry	Ferret	Church	Harp	Frog	Leaf	Snake	Ring	Hammer
1	Chain	Chair	Nose	Lettuce	Gun	Star	Motorcycle	Strawberry	Spider
1	Nest	Neck	Glove	Screw	Lightbulb	Pear	Snake	Rooster	Swing
1	Rattle	Radish	Helicopter	Zebra	Hammer	Button	Rooster	Truck	Brush
1	Candle	Candy	Television	Deer	Clown	Door	Cat	Tomato	Skunk
1	Beach	Beak	Watermelon	Ring	Pear	Clown	Football	Zebra	Screw
1	Cake	Cape	Piano	Bird	Doll	Thread	Clown	Bird	Tomato
1	Ball	Helicopter		Tiger	Glove	Screw	Mouse	Ring	Tomato
1	Train	Kangaroo		Nose	Bed	Fly	Deer	Doll	Ring
1	Chicken	Whistle		Zebra	Lightbulb	Train	Doll	Corn	Knife
1	Turtle	Ruler		Frog	Pear	Bed	Clown	Screw	Corn
1	Drum	Seal		Ring	Fence	Frog	Zebra	Vest	Book
1	Hammer	Spider		Cow	Pencil	Corn	Tomato	Book	Fox
1	Thread	Skunk		Flag	Corn	Lettuce	Hammer	Barn	Mouse
1	Key	Lips		Bird	Motorcycle	Lightbulb	Drum	Corn	Chain
1	Watch	Swing		Nail	Gun	Pear	Corn	Snake	Car
1	Fox	Fly		Snake	Button	Cow	Chain	Corn	Snake
2	Crater	Cradle	Key	Vest	Ruler	Key	Tomato	Mouse	Barn
2	Brick	Bridge	Swan	Chain	Strawberry	Watermelon	Bird	Lettuce	Doll
2	Rose	Rope	Lightbulb	Football	Deer	Bird	Rooster	Bird	Cow
2	Dime	Dice	Wrench	Truck	Cow	Deer	Knife	Hammer	Nail
2	Shade	Shark	Chicken	Lettuce	Frog	Strawberry	Corn	Piano	Deer
2	Goat	Goal	Piano	Doll	Spider	Pencil	Tomato	Truck	Ring
2	Bone	Bow	Cat	Ring	Leaf	Book	Vest	Chain	Pencil
2	Carrot	Carriage	Bird	Clown	Flag	Motorcycle	Nail	Lips	Snake
2	Horn	Horse	Giraffe	Drum	Seal	Doll	Bread	Pencil	Football
2	Bat	Bag	Shoe	Fox	Church	Screw	Ring	Spider	Lettuce
2	Grasshopper	Pipe		Spider	Strawberry	Pot	Harp	Clown	Lips
2	Cow	Bread		Glove	Leaf	Spider	Frog	Harp	Vest
2	Pot	Clown		Tomato	Tiger	Motorcycle	Swan	Harp	Swing
2	Harp	Pencil		Corn	Book	Tie	Bread	Lips	Bird
2	Nail	Lettuce		Swing	Thread	Key	Zebra	Sandwich	Bread
2	Piano	Seal		Cat	Motorcycle	Gun	Swing	Thread	Harp
2	Knife	Church		Lips	Fly	Book	Nail	Pear	Nose
2	Flag	Train		Sandwich	Screw	Car	Cow	Turtle	Ring
2	Wrench	Pear		Barn	Tie	Pipe	Bird	Bed	Drum
2	Fence	Watermelon		Bed	Turtle	Tiger	Wrench	Frog	Rooster
3	Saddle	Salmon	Tie	Rooster	Knife	Sandwich	Lettuce	Ring	Lips
3	Racket	Rabbit	Whistle	Clown	Glove	Hammer	Barn	Football	Truck
3	Hook	Hood	Kangaroo	Bird	Pipe	Fly	Book	Deer	Vest
3	Mouse	Mouth	Fly	Motorcycle	Key	Wrench	Football	Cow	Rooster
3	Mold	Mole	Plug	Skunk	Brush	Glove	Ring	Car	Swan
3	Shower	Shovel	Leaf	Piano	Ball	Church	Flag	Fox	Truck
3	Bell	Belt	Strawberry	Harp	Book	Tie	Vest	Nose	Cat
3	Money	Monkey	Drum	Tomato	Button	Clown	Skunk	Cat	Turtle
3	Drill	Drum	Watermelon	Lips	Bed	Gun	Pencil	Motorcycle	Deer
3	Mustard	Mustache	Deer	Spider	Tiger	Nose	Brush	Nail	Chain
3	Star	Deer		Strawberry	Screw	Tiger	Chicken	Football	Nail
3	Fox	Church		Chicken	Piano	Ruler	Deer	Drum	Motorcycle
3	Glove	Pot		Knife	Doll	Ruler	Sandwich	Wrench	Frog
3	Button	Whistle		Swan	Skunk	Doll	Corn	Flag	Football
3	Watermelon	Hammer		Harp	Clown	Peach	Swing	Spider	Bed
3	Tie	Vest		Thread	Door	Glove	Spider	Bread	Clown
3	Swing	Key		Pencil	Leaf	Brush	Bird	Glove	Bread
3	Gun	Fly		Bread	House	Leaf	Harp	Lettuce	Flag
3	Shoe	Mouse		Book	Tiger	Lightbulb	Bed	Harp	Bird
3	Grasshopper	Lightbulb		Rooster	Car	Tiger	Pear	Swing	Spider
4	Scale	Scarf	Rooster	Book	Star	Ball	Glove	Brush	Strawberry
4	Fork	Fort	Grasshopper	Truck	Clown	Flag	Lips	Tomato	Bird
4	Locket	Locker	Chicken	Ring	Key	Bed	Piano	Book	Thread
4	Cable	Caveman	Pipe	Hammer	Fly	Button	Turtle	Swing	Lettuce
4	Hand	Hat	Helicopter	Corn	Brush	House	Truck	Snake	Corn
4	Wine	Wire	Frog	Chain	Motorcycle	Book	Lettuce	Truck	Book
4	Toad	Toast	Peach	Football	Nose	Turtle	Truck	Zebra	Clown
4	Clown	Cloud	Drum	Swing	Gun	Skunk	Car	Nail	Truck
4	Pin	Pig	Church	Turtle	Bird	Seal	Chain	Bread	Glove
4	Basket	Badger	Shoe	Tomato	Tie	Pot	Swing	Clown	Pear
4	Skunk	Ball		Pear	Train	Frog	Ring	Swing	Piano
4	Hammer	House		Deer	Clown	Seal	Harp	Deer	Swing
4	Television	Piano		Brush	Watermelon	Clown	Spider	Bird	Tiger
4	Rooster	Watch		Vest	Peach	Leaf	Truck	Tomato	Harp

(*continued on next page*)

Table A2 (*continued*)

Set	Target	Prime	Alternate Prime	Copresent Filler 1	Copresent Filler 2	Copresent Filler 3	Separate Filler 1	Separate Filler 2	Separate Filler 3
4	Tomato	Barn		Nail	Ruler	Star	Nose	Chicken	Ring
4	Thread	Glove		Snake	Star	Clown	Ring	Chain	Harp
4	Door	Fox		Car	Book	Piano	Screw	Skunk	Zebra
4	Tiger	Cat		Truck	Wrench	Knife	Fox	Knife	Chicken
4	Football	Turtle		Bread	Lettuce	Brush	Book	Tiger	Sandwich
4	Flag	Key		Swing	Pot	Motorcycle	Tiger	Rooster	Corn

auditory memories during the planning process.

Data availability statement

Raw data, processed data, and experimental stimuli are available online at https://osf.io/ck2jg/.

Appendix A

See Tables A1 and A2.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2019.104066.

Acknowledgements

We sincerely thank Caoimhe Harrington Stack, Rachelle Brown, Ella Dixon, Josiah Pehrson, and Adrian Porras for their invaluable help in completing this project. The second author was supported by an NSF Graduate Research Fellowship. This project was funded in part by NSF-BCS 1557097 and the James S. McDonnell Foundation to the third author.

References

Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review, 19*, 505–512.
Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31–56.
Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.
Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*, 92–111.
Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer (Version 6.0.37). Retrieved from http://www.praat.org/.
Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*, 1044–1098.
Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1482–1493.
Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science, 1*, 274–291.
Brown, G. (1983). Prosodic structure and the given/new distinction. *Prosody: Models and measurements* (pp. 67–77). Berlin, Heidelberg: Springer.
Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes, 27*, 62–89.
Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. *Psychology of learning and motivation* (pp. 59–99). Academic Press.
Buxó-Lugo, A., Toscano, J. C., & Watson, D. G. (2018). Effects of participant engagement on prosodic prominence. *Discourse Processes, 55*, 305–323.
Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience, 31*, 404–424.
Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language, 89*, 68–86.
Bybee, J. L., & Hopper, P. J. (Vol. Eds.), (2001). *Frequency and the emergence of linguistic structure: Vol. 45*John Benjamins Publishing.
Clark, H. H. (1996). Communities, commonalities, and communication. In J. Gumperz, & S. Levinson (Eds.). *Rethinking linguistic relativity* (pp. 324–355). .
Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in psychology: Vol. 9*, (pp. 287–299). North-Holland.
Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology, 6*, 243–278.
Cohen Priva, U. (2017). Informativity and the actuation of lenition. *Language, 93*, 569–597.
Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327–339.
Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96*, 263–284.
Fisher, C., & Tokura, H. (1995). The given-new contract in speech to infants. *Journal of Memory and Language, 34*, 287–310.
Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504.
Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology, 19*, 509–525.
Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Vol. Eds.), *Syntax and semantics: Vol. 3, Speech Acts*, (pp. 41–58). New York: Academic Press.
Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning & Verbal Behavior, 13*, 512–521.
Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language, 47*, 589–606.
Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91–117.
Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language, 84*, 37–48.
Jescheniak, J. D., Kurtz, F., Schriefers, H., Günther, J., Klaus, J., & Mädebach, A. (2017). Words we do not say—Context effects on the phonological activation of lexical alternatives in speech production. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 1194–1206.
Jescheniak, J. D., Oppermann, F., Hantsch, A., Wagner, V., Mädebach, A., & Schriefers, H. (2009). Do perceived context pictures automatically activate their phonological code? *Experimental Psychology, 56*, 56–65.
Jescheniak, J. D., & Schriefers, H. (2001). Priming effects from phonologically related distractors in picture-word interference. *Quarterly Journal of Experimental Psychology, 54*, 371–382.
Kahn, J., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language, 67*, 311–325.
Kahn, J. M., & Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground. *Language, Cognition and Neuroscience, 30*, 103–119.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26.
Lam, T. Q., & Watson, D. G. (2014). Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory & Cognition, 40*, 829–843.
Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition, 38*, 1137–1146.
Levelt, W. J. M. (1989). *ACL-MIT Press series in natural-language processing. Speaking: From intention to articulation.* Cambridge, MA, US: The 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*, 1–75.
Lupyan, G., & Swingley, D. (2011). Self-directed speech affects visual processing.

Quarterly Journal of Experimental Psychology, 65, 1068–1085.
MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 226.
Mädebach, A., Kurtz, F., Schriefers, H., & Jescheniak, J. D. (2019). Pragmatic constraints do not prevent the co-activation of alternative names: Evidence from sequential naming tasks with one and two speakers. https://doi.org/10.31234/osf.io/9876v.
Mahrt, T. (2016). PraatIO. https://github.com/timmahrt/praatIO.
Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition, 35*, 494–503.
Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 555–563.
Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*, 528–537.
O'Seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 57.
Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language, 78*, 1–17.
Quine, W. V. (1968). Ontological relativity. *The Journal of Philosophy, 65*, 185–212.
R Core Team (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing https://www.R-project.org/.
Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika, 6*, 309–316.
Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*, 86–102.
Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53*, 91–127.
Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition, 133*, 140–155.
Tily, H., & Kuperman, V. (2012). Rational phonological lengthening in spoken Dutch. *The Journal of the Acoustical Society of America, 132*, 3935–3940.
Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes, 27*, 979–1001.
Watson, D. G., Buxó-Lugo, A., & Simmons, D. C. (2015). The effect of phonological encoding on word duration: Selection takes time. In E. Gibson, & L. Frazier (Eds.). *Explicit and implicit prosody in sentence processing* (pp. 85–98). Switzerland: Springer International Publishing.
Yiu, L. K., & Watson, D. G. (2015). When overlap leads to competition: Effects of phonological encoding on word duration. *Psychonomic Bulletin & Review, 22*, 1701–1708.
Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 919–937.
Zerkle, S., & Arnold, J. E. (2016). Discourse attention during utterance planning affects referential form choice. *Linguistics Vanguard, 2*.