VARIATION IN WORD DURATION AND PLANNING

Alexandros Christodoulou

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology.

Chapel Hill
2012

Approved by:

Jennifer E. Arnold

Patrick J. Curran

Peter C. Gordon

Neil W. Mulligan

Jennifer L. Smith

ABSTRACT

ALEXANDROS CHRISTODOULOU: VARIATION IN WORD DURATION AND PLANNING
(Under the direction of Jennifer E. Arnold)

Word duration varies as a function of predictability (e.g., Bell et al., 2009; Lieberman, 1963; Fowler & Housum, 1987; Gahl & Garnsey, 2004; Watson, Arnold, & Tanenhaus, 2008) and planning difficulty (e.g., Bell et al., 2003; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). Both findings suggest that planning is related to word duration variation. However, the role of planning in word duration variation is debated (e.g., Arnold & Watson, under review; Ferreira, 2007). In four experiments I tested the hypotheses that planning an upcoming word and variation in the timing of planning an upcoming word leads to word duration variation of an utterance-initial word. Speakers named two pictures while trying not to pause. The utterance-initial word was long in duration when followed by a low frequency word. Furthermore, looks to the right object modulated the effect of frequency on word duration. Early looks led to word reduction when the second word was high frequency. Late looks led to long word duration regardless of frequency. The results support the role of planning on word duration. Theories of utterance planning should consider evidence of word duration variation apart from the traditional speech onset time and utterance duration measures.

DEDICATION

To my loving parents Andreas and Maria Christodoulou.

ACKNOWLEDGEMENTS

I would like to thank my advisor Jennifer E. Arnold for her continuous intellectual support throughout my years at UNC Chapel Hill. I would also like to thank the members of my dissertation committee Patrick J. Curran, Peter C. Gordon, Neil W. Mulligan, and Jennifer L. Smith. They gave me invaluable feedback, alongside my advisor, that contributed to the success of my dissertation project.

I am also very thankful to my friends and colleagues in the psychology department at UNC. Miri, Ned, So-Yeon, and Vicki made the psychology program a very happy workplace for me. Of course, this is when I was not organizing the prospective students' visiting day and being yelled at for my choice of restaurants. Also, I am grateful to my lab-mates Jason and Elise for providing some great feedback at different stages of my dissertation project.

Last but not least, I would like to thank my parents for giving me the opportunity to pursue my dream of studying psychology in the US far away from my home in Greece. I hope I have made them happy and also proud of me.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

Utterance planning can influence word duration variation. For example, imagine that So-Yeon is hungry and goes to order a sandwich. At the sandwich store the ingredients are behind a glass display. As she scans the ingredients' display she sees olives. Because she usually orders tomatoes, which are usually located next to the olives, she decides to begin speaking. Right before she starts articulating *olives* she looks next to the olives but unfortunately she finds pickles. She decides to prolong the articulation of *o..l..i..v..e..s..* to look next to the pickles. At the onset of *o..l..i..v..e..s..* she is relieved to find tomatoes. In case she cannot easily find the tomatoes, she could insert the conjunction *and* to allow additional time to find them in the display. Furthermore, she should also abide by social conversational rules of being brief and fluent (Clark, 2002; Clark & Wasow, 1998; Grice, 1975). Under these conditions, will there be evidence that variables that influence utterance planning, such as the timing and difficulty of planning words upcoming in the discourse, influence word duration?

Utterance planning occurs at two separate stages, the grammatical encoding stage, and the phonological encoding stage (Bock & Levelt, 1994; Levelt, 1989). Planning at these two stages occurs after an intended message has been generated. In this dissertation I will consider this stage as part of utterance planning because message generation interacts with utterance planning (Brown-Schmidt & Tanenhaus, 2006; Brown-Schmidt & Konopka, 2008). In the previous example about ordering a sandwich, the generation of the concept to order olives and tomatoes corresponds to the message generation. Retrieval of the target words followed by word order determination and the appropriate suffix addition (i.e. the addition of the plural –s) corresponds to the grammatical encoding stage. Retrieval of instructions about how each word should sound corresponds to the phonological encoding stage.

Utterances can be considered preplanned when all phonological information of the utterance is planned prior to speech onset (Sternberg, Monsell, Knoll, & Wright, 1978; Sternberg, Wright, & Monsell, 1980;). Utterance planning can also occur incrementally when speakers plan only a minimal unit of grammatical or phonological information prior to speech onset (Griffin, 2001; Kempen & Hoenkamp, 1987; Levelt & Meyer, 2000; Meyer, 1996; Meyer, Sleiderink, & Levelt, 1998; Schriefers & Teruel, 1999; Wheeldon & Lahiri, 1997). There is evidence that rehearsed utterances can be preplanned whereas extemporaneous speech is usually planned incrementally (Sternberg et al., 1978, 1980; Wheeldon & Lahiri, 1997).

Furthermore, there is a debate on the unit of grammatical and phonological planning. The length of the unit i.e. a syllable or a word defines the scope of planning or how much information is planned prior to speech onset (Allum & Wheeldon, 2007; 2009; Levelt & Meyer, 2000; Meyer, 1996; Meyer et al., 1998; Oppermann, Jescheniak, & Schriefers, 2010; Schnur, Costa, & Caramazza, 2006; Schriefers & Teruel, 1999; Wagner, Jescheniak, & Schriefers, 2010; Wheeldon & Lahiri, 1997). Specifically for phonological planning some evidence suggests that the unit can be as small as a syllable (Schriefers & Teruel, 1999; Meyer, Belke, Haecker, & Mortensen, 2007) or a phonological word i.e. a content word and a function word (Wheeldon & Lahiri, 1997).

Other evidence suggests that it can be as large as three phonological words (Oppermann, et al., 2010; Schnur, et al., 2006). Therefore the unit of phonological planning, or the scope of planning, is variable. This suggests that the amount of incremental planning relative to the amount of preplanning can vary.

The effect of utterance planning on word duration is supported by findings that words are long before disfluent pauses (Bell et al., 2003; Cark & Fox Tree, 2002; Fox Tree & Clark, 1997; Shriberg, 2001). Disfluent pauses are non-grammatically motivated pauses that tend to have long duration e.g. longer than 200 ms or 250 ms (Goldman Eisler, 1968; Levelt, 1989). Pausing and the associated long word duration therefore occur because of planning. The effect of utterance planning on word duration is also supported by findings that words are long when utterance planning is difficult e.g. when the speaker is describing an unfamiliar object as in figure 1 (Christodoulou, 2009). In both phenomena it is reasonable to assume that speakers try to create additional planning time by extending the duration of their words.

Even though there is evidence that utterance planning is related to word duration variation, there is yet no direct evidence that variation in the incrementality of planning is also related to word duration variation. Ferreira & Swets (2002) provide evidence that utterance duration varies when speakers switch from preplanning to incremental planning. Speakers calculated arithmetic sums and stated the sum using the following phrase *the sum is.* Speech onset time was short but the duration of the phrase was long when the sum was difficult to calculate and when speakers were under a deadline to begin speaking. However from this finding it is not clear which words in the phrase varied in duration and whether pausing drove the utterance duration effect.

Overall planning strategy of preplanning vs. incremental planning (or the relative degree of incrementality) can be detected by using a global measure such as utterance duration in addition to the measure of speech onset time as in Ferreira and Swets (2002). Changes in speech onset time relative to changes in utterance duration should index changes in the amount of information preplanned relative to the amount of information planned incrementally. These measures however cannot accurately index processing time of a specific upcoming word in an utterance. For example, in the simple utterance *olives and tomatoes,* speech onset time and total utterance duration cannot provide an accurate measure of the timing of the second word planning. Such a measure can be informative about the relationship between planning and word duration given evidence that the probability of a following or a preceding word can influence the current word duration (e.g., Bell et al., 2003, 2009; Jurafsky, Bell, Gregory, & Raymond, 2001).

Griffin (2003) provides evidence that the timing of planning in a two-word utterance relative to speech onset can vary according to the duration of the utterance-initial word. For example, speech onset time is short when the two-word sequence is *windmill carrot* rather than *wig carrot.* The timing of planning can be defined as the difference between speech onset time and when information about a particular word in an utterance has been significantly processed, such as the onset of the last fixation to the object to be named (e.g. Griffin & Bock, 2000). This is also called the eye-voice span.

The eye-voice span should index visual processing of an object to be named, as well as lexical processing time for the object (Griffin, 2001; Meyer & van der Meulen, 2000; Meyer et al., 1998, 2003, 2007). Processing could occur either before or after speech onset. However, it is an imperfect measure of lexical processing because as

Morgan and Meyer (2005) suggest, speakers tend to look to the next object prior to complete processing of the previous word. Speakers tend to shift their gaze to the next object when the current word has been phonologically planned (Meyer & van der Meulen, 2000; Meyer et al., 1998, 2003, 2007). Therefore, this measure is suitable for answering questions about how much information up to phonological encoding has been planned prior to referring to the target object.

I am interested in isolating the point in time at which processing of the second object begins, even if it occurs in parallel with some residual processing of the first object. Therefore I will define the eye-voice span as the difference between the onset of the first fixation to a second object and speech onset time. This measure should index the onset of processing the second object. In comparison, Griffin and Bock's (2000) eye-voice span definition is the difference between the onset of the last fixation at the target object and the onset of the word referring to that object. I will return to the issue of parallel processing in the general discussion.

The combination of evidence of variation in incrementality (Ferreira & Swets, 2002; Griffin, 2003) with the effect of predictability on word duration (Bell et al., 2003, 2009; Jurafsky et al., 2001) might suggest a direct relationship between variation in incrementality and variation in word duration. Their relationship could suggest that slight adjustments to the timing of planning have measurable consequences on word duration. These adjustments could occur if speakers use flexible rules about speech initiation.

For example on one occasion the speaker might see the tomatoes and decide to initiate speech after phonologically encoding part of the word. In a different occasion the speaker might see the tomatoes and decide to initiate speech without phonologically encoding the word. This could potentially occur if the speaker starts speaking betting they will find the tomatoes and plan the word while articulating the first word.

This example leads to one of the main research questions addressed in this dissertation, which is whether the timing of planning can affect word duration. If speakers modulate word duration to accommodate planning needs, word duration could be long when they have not already begun processing the second word. *Olives* could therefore vary in duration because of variation in the speed of accessing visual, grammatical or phonological information of *artichokes* or *tomatoes*. Thus, my first hypothesis of how planning can affect word duration is that variation in the timing of planning a word upcoming in the utterance such as *artichokes* or *tomatoes* affects the duration of an utterance-initial word such as *olives*.

The other main research question is whether word duration varies because of planning difficulty of upcoming words. *Artichoke* is lower in frequency compared to *tomato* and should be slower to access in the mental lexicon (e.g., Oldfield & Wingfield, 1965). Therefore, *olives* could be longer in duration when followed by *artichokes* compared to *tomatoes,* if speakers plan while speaking and if they modulate word duration in order to accommodate planning needs. On the other hand, if planning is unrelated to word duration, the duration of *olives* should be unchanged in the previous two scenarios keeping other factors constant such as rate of speech.

The second hypothesis of how planning can affect word duration is that the long duration of planning associated with a low frequency word such as *artichoke* can lead to long duration of *olives.* If speakers modulate word duration because of planning an upcoming word, an upcoming word that requires long planning time such as *artichoke* should lead to long word duration compared to an upcoming word that requires short planning time such as *tomato*.

This process on its own could affect the duration of *olives* or it could be modulated by the variation in the second word timing of planning. Variation in the timing of planning could override any effects of planning difficulty by making both high and low frequency words easy or difficult to plan together with the first word. For example if the artichokes or tomatoes are located next to the olives, speakers could preplan part or all of *artichokes* or *tomatoes* prior to speech onset. This should reduce the need to modulate word duration because of planning. Otherwise, the need to modulate word duration because of planning could be minimized only for words that do not require long planning time such as *tomatoes*. Therefore, a low frequency upcoming word should lead to long utterance-initial word duration independent of the timing of planning.

The duration of utterance planning or timing of planning variation could also lead to pausing. For example Goldman Eisler (1968) has shown that conceptual complexity leads to more frequent pausing. Even though this effect is theoretically interesting on its own, I will not focus on it. To minimize the degree of pausing in the reported experiments I will ask speakers to try not to pause when naming consecutive objects (e.g., Griffin, 2003).

The purpose of this dissertation is to test the hypothesis that planning affects word duration using the dual-picture naming task (e.g., Meyer, 1996). The hypothesis aims to distinguish between different proposals of whether planning should be related to word duration variation (e.g., Arnold & Watson, under review; Ferreira, 2007). The secondary purpose of this thesis is to understand how planning is related to word duration. I address this by testing two separate hypotheses: (1) planning a difficult upcoming word leads to long word duration compared to planning an easy upcoming word (2) the timing of planning the upcoming word leads to word duration variation.

In the following sections I provide evidence for word duration variation and for variation in the amount and timing of planning. I then describe predictions about how planning can affect word duration. I close the introductory chapter with the experimental paradigm that will allow the test of the relationship between planning and word duration.

Variation in Word Duration

Word duration varies in conversational speech (e.g., Bell et al., 2009) and part of this variation is possibly because of planning demands (e.g. Bell et al., 2003). However, there is little work on how utterance planning can affect word duration. This is probably because word duration is sensitive to multiple factors that co-occur in natural speech such as the presence of syntactic or prosodic boundaries (e.g., Ferreira 1993) and the rate of speech (e.g., Fosler-Lussier, Morgan, 1999).

Word duration can vary in different ways. This is exemplified in the phenomenon of reduction. Reduction can occur in different forms: reduction in acoustic duration, changes in vowel quality (e.g., the definite article can have two variants *thee* and *thu*), segment deletion (e.g., in the word *confirmed* the *d* sound is not pronounced). In this dissertation I will only measure acoustic duration because I am interested in word duration variation and not just reduction.

A well established phenomenon related to word reduction is Givenness (e.g., Fowler & Housum, 1987). Words or their referents that are given in the discourse tend to be reduced in duration. This phenomenon is one of the different discourse properties that can influence the accessibility of information (i.e., ease of accessing information relevant to the discourse; Arnold, 2010). For example in the following hypothetical discourse segment:

I ate a sandwich

The sandwich had olives

The second mention of *sandwich* will be reduced in duration because its referent and lexical form have already been presented in the discourse (Balota, Boland & Shields, 1989; Bard et al., 2000; Fowler & Housum, 1987).

Word duration is inherently variable. In certain situations a word can be short as in the case of reduction. However, in other situations a word can be long as in the case of the presence of a following disfluency (e.g., Bell et al., 2003). Therefore, the reduction literature needs to be complemented by the disfluency literature (and planning literature) in order to fully understand why word duration varies because of processing demands. Here I will refer to two separate phenomena, reduction and lengthening, only because of how the different literatures have categorized word duration as either reduced or lengthened. However, I will not assume that reduction is a different phenomenon than lengthening, rather that word duration varies between reduced and lengthened values.

Reduction is related to the context in which a word appears. Lieberman (1963) proposed that the probability of a word given the preceding semantic context leads to word duration variation. For example, the word *nine* is more probable and thus reduced in duration in the sentence: *"a stitch in time saves __"* compared to the sentence: *"the time is__".* Since Lieberman's (1963) work, predictability has been recast in different ways. One line of research treats predictability as the probability of a word or syllable appearing within the context of another word or syllable (Aylett & Turk, 2004; Bell et al., 1999, 2002, 2003, 2009; Jurafsky et al., 2001, 2002).

There are currently two prominent hypotheses about how predictability relates to reduction. According to the probabilistic reduction hypothesis (PRH) (e.g. Jurafsky et al., 2001) words are reduced according to the conditional probability of a given word. For example in the two-word sequence, *Harrison Ford*, *Ford* is highly predictable given the word *Harrison* (i.e., if I just heard *Harrison* there is a good chance it will be followed by *Ford*). Similarly, in the two-word sequence *Burt Reynolds*, *Burt* is highly predictable given the word *Reynolds* (i.e., if I just heard *Reynolds* there is a good chance it was preceded by *Burt*).

According to the smooth signal redundancy hypothesis (SSRH) speakers produce utterances "whose elements have similar probabilities of recognition" (Aylett & Turk, 2004). This is a similar concept to conditional probability in that a word or syllable is likely to be recognized when it is predictable given the preceding or following word or syllable. Probable words or syllables are then reduced because the listener needs only minimal amount of exposure to successfully recognize them. For example, in the three-syllable sequence *Burt-Rey-nolds*, the syllable *nolds* will be reduced because it will be easy for a listener to recognize given the preceding two syllables. Therefore both the SSRH and the PRH make the same prediction about the relationship between information redundancy and word duration. Redundant words given the context tend to be reduced.

According to both the PRH and SSRH the source of reduction is at the lexical or sublexical level. However, words are embedded in superordinate linguistic structures such as syntactic structures. A different line of research argues that predictability measures that lead to word duration variation should also include measures at the level of syntax (Gahl & Garnsey, 2004; Gahl & Garnsey, 2006; Gahl, Garnsey, Fisher, & Matzen, 2006; Tily et al., 2009). Verb bias is the probability of a verb appearing in a specific syntactic context. For example, the verb *confirm* is usually followed by a direct object (see the following example (1)) as opposed to a sentential complement (see the following example (2)).

(1) The CIA director confirmed the rumor once it had spread widely.

(2) The CIA director confirmed the rumor should have been stopped sooner.

Gahl & Garnsey (2004) showed that verbs followed by a complement matching their bias have a higher chance of undergoing final –t/-d deletion.

However, there is also evidence that the source of reduction can occur at even higher levels of linguistic planning, the discourse level. At this level, speakers integrate the current utterance with the context of the previous and potentially following utterance(s). The most common phenomenon at the discourse level is that words are reduced when they, or their referents, are given in the discourse (Bard & Aylett, 1999; Bard et al, 2000; Fowler, 1988; Fowler & Housum, 1987; Fowler, Levy, & Brown, 1997). For example, Fowler and Housum (1987) analyzed recorded speech from six different radio shows (a monologue and five interviews) and found that words (e.g., the word *match* when referring to a *tennis match*) that appeared at least twice in the discourse were shorter compared to their first mention.

Furthermore, a referent that is predictable independent of the discourse history can lead to duration variation. Watson, Arnold, & Tanenhaus (2008) analyzed utterances of game moves in a game of Tic-Tac-Toe. Words referring to a predictable location such as blocking a winning move were reduced compared to words referring to unpredictable locations.

Another phenomenon that has been studied with relation to word duration variation is lengthening. Words are long in duration when followed by a disfluent pause (i.e., a non syntactically motivated pause) and when the speaker encounters conceptual difficulty. One of the best examples of how disfluent pausing is related to word duration variation comes from disfluent repetitions (Shriberg, 1999, 2001). For example, when a speaker tries to describe a conceptually unfamiliar image as the one in figure 1 they might say something like:

the(R1) …. *the*(R2) *blue sunset over a lake.*

The first mention of *the*(R1) is prolonged compared to the second mention of *the*(R2). The effect of lengthening on word duration generalizes to all types of lexical disfluencies such as when speakers insert words like *um* or *uh* in the utterance (Bell et al., 2003; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997).

Evidence from a different line of experimentation suggests that task conceptual difficulty can influence word duration (Christodoulou, 2009). For example, I asked speakers to describe unfamiliar pictures such as in figure 1 using the following carrier phrase: *click on the (color* of object). I found that conceptually unfamiliar pictures led to longer color word duration compared to familiar pictures.

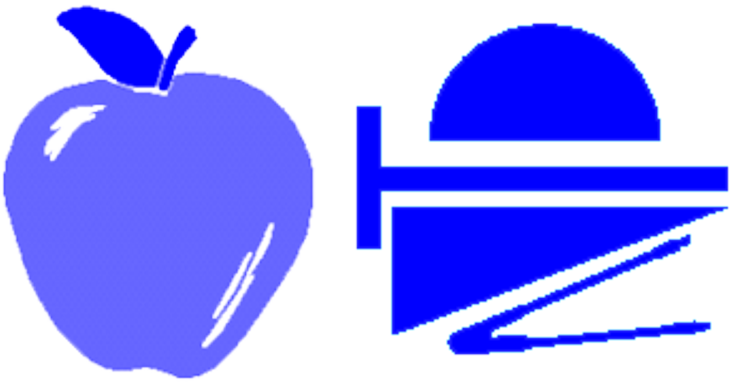


Figure 1. Example of a familiar vs. unfamiliar picture from Christodoulou (2009). The picture next to the apple was commonly referred to as the sunset over a lake.

The combination of evidence from the reduction and lengthening literature suggests that planning can affect word duration variation. The effect of predictability on word duration could be mediated by planning. Even though a word might be highly probable given the lexical, syntactic, or discourse context; the word, the context, and how the word fits into the context has to be planned on line. For example, when the speaker plans the two-word sequence *Burt Reynolds,* the duration of *Burt* will be reduced potentially because the two words are easy to plan together. In this example, this would entail planning *Reynolds* before the onset of *Burt.*

Both disfluent contexts and cases of conceptual difficulty are related to lexical planning difficulty. In both cases it is reasonable to assume that speakers are slowing down their rate of articulation because upcoming information is difficult to retrieve.

Therefore, by slowing down the rate of articulation the speaker tries to buy additional planning time.

In sum, there is evidence to suggest that planning affects word duration. However, there is variation in the timing of planning (e.g., Griffin, 2003). This raises the question of whether word duration variation is sensitive to differences in the timing of planning. In the following section, I describe evidence of scope of planning variation and timing of planning variation and how it could be related to word duration.

Variation in Planning

The timing of planning varies (Griffin, 2003; Meyer, 1996; Schriefers & Teruel, 1999; Wagner et al., 2010). There is however limited evidence that this variation is related to word duration variation. One possibility for the limited evidence is that research on utterance planning has mainly focused on traditional measures of speech onset time and utterance duration (e.g., Ferreira & Swets, 2002; Meyer et al., 2007; Schriefers & Teruel, 1999; Wheeldon & Lahiri, 1997). In a multiword utterance speech onset time may not be sensitive to minor changes in planning demands. Furthermore, even though utterance duration should be correlated with word duration it is confounded with pausing.

In this section I present evidence about the possible relationship between variation in the timing of planning and word duration. I also highlight some factors that need to be controlled to draw conclusions about their relationship. These factors include: utterance length (Sternberg et al., 1978, 1980; Griffin, 2003; Meyer, 1996), type of naming task (e.g., Griffin, 2003; Meyer, 1996), control for when to plan (Ferreira &

Swets, 2002; Griffin, 2003), and speaker characteristics (Schriefers & Teruel, 1999; Wagner et al., 2010).

The timing of planning upcoming information is in part determined by decisions about articulation initiation. Speakers can use different criteria about when to initiate articulation. For example, in the minimal utterance *olives and tomatoes*, speech onset could occur after preparing the syntactic or phonological information of the first word, as in incremental planning (e.g., Griffin, 2001; Kempen & Hoenkamp, 1987; Levelt, 1989; Levelt & Meyer, 2000; Meyer, 1996; Meyer et al., 1998). It could also occur when phonological information of up to all words in the utterance has been prepared, as in preplanning (Goldman Eisler, 1968; Sternberg et al., 1978, 1980).

Figure 2 contrasts incremental planning (top panel) with preplanning (bottom panel) in a dual-picture naming task. The difference between the two types of planning can be seen as two ends on a continuum. However, there also needs to be categorical distinction between the two given evidence that speakers can switch between the two different types of planning (Ferreira & Swets, 2002; Wagner et al., 2010).

The two pictures at the top of the figures depict the objects to be named and the red dashed line the onset of articulation. In both figures each rectangle represents a different planning stage according to dominant models of utterance and lexical planning (Bock & Levelt, 1994; Dell, 1986; Levelt, Roelofs, Meyer, 1999). The depicted stages are similar to the described planning stages according to Bock & Levelt (1994). First the speaker visually inspects the target object, which leads to accessing conceptual information about the object. The conceptual information is then semantically and syntactically encoded before being phonologically encoded. At the final stage the planning system assembles and starts executing the motor commands for the production of the target words.

According to incremental planning, speech onset time occurs earlier compared to preplanning. This is because speakers can start articulating the first word as they retrieve semantic/ syntactic information of the second word. However, according to preplanning, speech onset is delayed until all linguistic information of both words is retrieved.

We can observe variation in the timing of planning by comparing tasks that encourage incremental planning (Meyer, 1996; Schriefers & Teruel, 1999) with those that encourage preplanning (Alario, Costa, & Caramazza, 2002; Costa & Caramazza, 2002; Schnur et al., 2006; Sternberg et al., 1978; 1980). We can also observe variation when comparing speakers with different characteristics within a single task, such as error rate in naming accuracy and speech rate (Schriefers & Teruel, 1999; Wagner et al., 2010). Furthermore, some speakers can change their criterion of when to initiate articulation and therefore switch from incremental planning to preplanning and vice versa. This change can be induced by a deadline to begin speaking (Ferreira & Swets, 2002) or by manipulating cognitive load (Wagner et al., 2010).

There is evidence that speakers can initiate articulation only after preparing phonological information of a single word prior to speech onset. For example, Meyer (1996) in experiment 3 asked speakers to name two pictures using noun phrase conjunctions such as *the bag and the arrow* while ignoring a superimposed interfering word. The word was phonologically related or unrelated to the name of the first or the second picture. An example picture-word pair would be *bag – bat*. The interfering

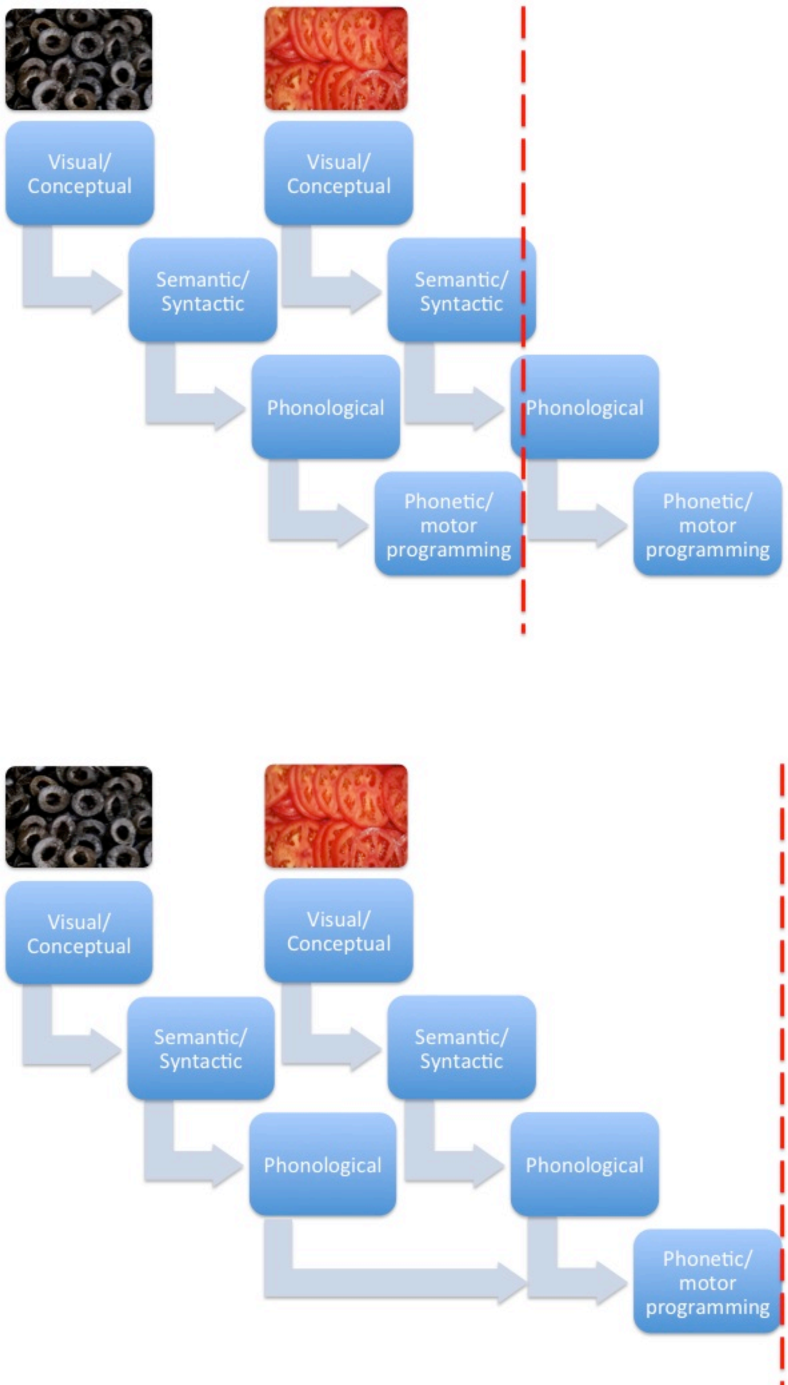


Figure 2. Example of incremental planning vs. preplanning. The pictures at the top of the diagram depict two example stimuli pictures from a hypothetical dual-picture naming task. Each rectangle depicts a different planning stage. The red dashed line depicts the onset of speech

word decreased speech onset time only when it was related to the first word but not the second word.

This finding suggests that the phonological information of only the first word was available prior to speech onset. In a separate experiment Meyer (1996) found that semantic information of both words was available prior to speech onset. The combination of the two findings by Meyer (1996) suggests that the timing of planning the second word differs for phonological compared to semantic information. This difference can be interpreted according to an incremental planning account. The speaker starts planning phonological information of the first word after accessing syntactic information of the second word. They then plan phonological information of the second word after accessing syntactic information of the first word.

On the other hand, other evidence suggests that sometimes speakers can initiate articulation after preparing two phonological words (roughly a function word and a content word) (Alario et al., 2002; Costa & Caramazza, 2002; Cholin, Dell, & Levelt, 2011; Schnur et al., 2006; Schriefers & Teruel, 1999). For example Schnur et al. (2006) asked participants to describe actions performed in single pictures e.g., *the orange girl jumps* while ignoring a phonologically related word. This utterance contains three phonological words (1. *The orange,* 2. *girl* 3. *jumps)*. Their results showed that speech onset time was significantly reduced even for phonological distractors related to the verb. As Schnur et al. (2006) suggests, the difference in the degree of incremental planning between tasks could be related to conceptual integration. *The orange girl jumps* refers to a single entity whereas *the bag and the arrow* refers to two distinct entities that are not strongly conceptually related in the utterance.

Therefore, the findings so far suggest that speakers can vary the amount of preplanned information according to task demands. If we assume that planning has a higher chance of affecting word duration when speakers plan as they speak, it is therefore preferable to use a dual-picture naming task as opposed to a single-picture naming task because it encourages incremental planning.

However, apart from task demands, speaker internal characteristics such as cognitive load can lead to variation in the timing of planning. Wagner et al. (2010) manipulated cognitive load in a dual-picture naming task as in Meyer (1996). In two experiments they varied syntactic complexity as follows:

(1) The frog is next to the mug (simple sentence)

(2) The red frog is next to the red mug (complex sentence)

In two additional experiments they combined picture naming with a secondary task. In one experiment the speakers learned to pair the outcome of a conceptual categorization task with either of the two syntactic forms used in the previous experiments. For example, if the frog and the mug fit into a drawer, the speakers were instructed to produce a simple sentence as opposed to a complex sentence. In the last experiment speakers were required to memorize five words or digits before stimulus onset. Upon utterance production they were given a recognition task of the previously memorized list.

The degree of incrementality varied according to syntactic complexity. Complex sentences increased the interference effect for the first noun. The interference effect was even stronger when syntactic complexity was paired with the categorization task. Interestingly, when syntactic complexity was paired with the span task the interference effect for the second noun increased while the interference effect of the first noun remained stable. The authors therefore concluded that the degree of incrementality is flexible because of differences in cognitive load. This raises questions about when the speaker encounters cognitive load in other production tasks and how it can affect the timing of planning. Planning two difficult words together could increase cognitive load, which in turn could affect the timing of planning the second word.

The word duration variation literature suggests that word duration is long when the speaker encounters planning difficulty (e.g., Christodoulou, 2009). This could potentially occur because the speaker adopts a higher degree of incremental planning under cognitive load as Wagner et al. (2010) suggests. Therefore, one prediction is that delayed timing of planning the second word in a two-word utterance, with relation to speech onset, leads to long word duration.

Schriefers & Teruel (1999) categorized speakers as having high vs. low error rate/pausing rate. Speakers with low error rate had a lower degree of incremental planning and therefore preplanned more than speakers with high error rate. Similarly, Wagner et al. (2010) categorized speakers according to speech onset time into slow vs. fast speakers. Fast speakers showed a higher degree of incremental planning compared to slow speakers. Therefore, both of these studies suggest that the degree of incremental planning varies. This variation can come from speaker characteristics of how they produce their utterance.

These results complicate the possible relationship between variation in the timing of planning and word duration. On the one hand, assuming that high error rate in naming is associated with cognitive load, word duration could be long the less speakers preplan, as the combination of the Wagner et al. (2010) and Christodoulou (2009) results suggest. On the other hand, if fast speech rate is associated with short word duration the Wagner et al. (2010) results might suggest that word duration could be short when speakers preplan less. Therefore, speaker characteristics can potentially provide competing forces on word duration. It is therefore important to control these characteristics. This can be implemented by using minimal utterances such as in the dual-picture naming task used by Griffin (2003) and by statistically controlling for speaker characteristics via multilevel modeling (e.g. Baayen, Davidson, Bates, 2008).

Furthermore, task constrains such as a deadline to begin speaking can lead to variation in the degree of incremental planning. In Ferreira & Swets (2002) speakers computed arithmetic sums. According to one condition they reported the sum by using the carrier phrase “the sum is XX”. In a second experiment speakers were required to initiate speech before a specific deadline. Speakers preplanned less but planned more incrementally when facing a deadline to begin speaking. Speech onset time was overall reduced but utterance duration was overall lengthened when the speaker faced a deadline to begin speaking.

Therefore the Ferreira & Swets (2002) results suggest a relationship between variation in the degree of incremental planning and utterance duration. Their results also imply an interaction between variation in incrementality and planning difficulty. In the first experiment the difficulty of computing the arithmetic sum did not influence the duration of *the sum is*. However, in experiment 2 the same segment was long when comparing the more difficult two-addend problems compared to the easier mixed-addend problems (e.g., 66 + 21 is more difficult compared to 2 + 41).

The previous evidence suggests a possible relationship between variation in the timing of planning an upcoming word and variation in word duration but only indirectly. Evidence by Griffin (2003) suggests a direct relationship between variation in the timing of planning and variation in word duration. In a dual-picture naming task, Griffin (2003) asked speakers to produce bare nouns such as *windmill carrot*. Speech onset time, and the duration of looks to the right object, were reduced when Word1 was multisyllabic such as the word *windmill,* rather than monosyllabic such as the word *wig.* This reverse length effect disappeared when the speaker introduced additional words between the two bare nouns such as *next to*.

Griffin (2003) suggests that when Word1 is monosyllabic, the speaker does more preplanning of the whole utterance, because the monosyllabic word does not afford incremental planning time (i.e., planning while speaking). However, when Word1 is multisyllabic the speaker does less preplanning of the whole utterance because the multisyllabic word affords enough incremental planning time to plan some of Word2 while articulating Word1. Griffin (2003) attributes this ability of adjusting the timing of Word2 planning to the speaker's ability to estimate Word1 duration.

Meyer et al. (2007) proposes a different mechanism that can explain Griffin's (2003) results. They argue that speakers plan incrementally when they can set a self-imposed deadline of when to start speaking. Speakers prefer to initiate speech after processing a single syllable. Instead of presenting all words in a single mixed list, they presented monosyllabic words in a separate block from trisyllabic words. They found no speech onset difference according to block but they still found an eye-voice span difference (i.e. the difference between the last fixation to the left object and speech onset). In a different experiment they replaced the second word with a symbol recognition task, and replicated their results. The authors argue that these results are compatible with a radical incremental account where speakers can initiate speech after preparing only a single syllable. The difference in the timing of planning (as measured with the eye-voice span between the last fixation to the left object and the onset of speech) occurs because in monosyllabic words the speaker has to look at the right object after processing a single syllable, which is also the end of the word. In trisyllabic words the speaker does not have to look at the right object until they have processed the third syllable.

Both Griffin (2003) and Meyer et al. (2007) provide evidence that utterance-initial word duration can provide planning time for the upcoming word. Furthermore, Griffin (2003) argues that people make decisions about when to begin speaking based on their assumptions about how much planning time they need for Word2 and how much time they will have to plan incrementally. Importantly, both of these studies suggest that manipulated word length is associated with the timing of planning without requiring variation in the scope of planning. Griffin (2003) argues that the timing of planning varies because the speaker estimates the amount of available planning time for an upcoming word. On the other hand, Meyer et al. (2007) argues that the timing of planning varies because of the first word syllable length and independent of planning an upcoming word.

Importantly, as Meyer et al (2007) argue, the two accounts are complementary. This raises questions as to whether variation in the timing of planning can lead to word duration variation and whether the variation in word duration could occur because of planning the first word or the second word in a two word utterance. This question merits further investigation in the face of evidence from the predictability literature. The predictability of a word given a preceding or following word can lead to reduction (e.g., Bell et al., 2003, 2009; Jurafsky et al., 2001). Reduction could occur because predictability shifts the timing of planning the preceding or upcoming word relative to the current word. In this dissertation I will focus on how the timing of the second word can influence the duration of the first word, in a dual-picture naming task. I will accomplish this by controlling the number of syllables and planning difficulty of the first word and manipulating planning difficulty of the second word.

Overall, evidence in this section suggests an inverse relationship between variation in the timing of planning a word (or words) upcoming in the utterance and word duration when a series of controls are used.

Predictions about Planning Effects on Word Duration

There are two views based on the literature that make opposite predictions about the role of planning on word duration. One view suggests that planning affects variation in word duration (Arnold & Watson, under review; Balota, et al., 1989; Kello, Plaut, & MacWhinney, 2000). The other view suggests that planning is unrelated to word duration variation (Damian, 2003; Ferreira, 1991, 1993, 2007).

There are no explicit proposals about the relationship between variation in the timing of planning and word duration. However, there is evidence for an inverse relationship between variation in the timing of planning an upcoming word and the current word duration. Early timing of planning should lead to word reduction. This prediction is based on the combination of evidence from the timing of planning variation and word duration variation literature. Nevertheless, this inverse relationship might occur only under specific circumstances. These circumstances might include the requirement to either not pause (Griffin, 2003) or the requirement to initiate speech before a certain deadline (Ferreira & Swets, 2002).

A proposal about the role of planning on word duration variation is put forth by Arnold & Watson (under review). They argue that the source of word duration variation is often ambiguous. For example, the discourse structure can guide decisions about accenting new and deaccenting old information. Accenting is associated with word duration variation but also with pitch and intensity variation. New information tends to be long in duration but old information tends to be reduced (e.g., Fowler & Housum, 1987). However, the source of the reduction is either guided by a linguistic rule that requires old information to be deaccented, or by the effect of planning facilitation. Repeated words are primed and therefore easier to plan (e.g., Balota et al., 1987).

The view that planning facilitation is related to word duration variation is supported by Balota et al. (1989). They used a priming task where a target word was either preceded or followed by a conceptually related word e.g., *cat … dog*. The target word was significantly reduced when paired with a related vs. an unrelated prime.

Furthermore, planning inhibition leads to relatively longer pronunciations. Inhibited words tend to be difficult to plan because their semantic or phonological information requires additional time to become available. For example, Kello et al. (2000) found evidence that word duration varies in a word interference task (i.e. a Stroop like task) when speakers have to start speaking before a specific deadline. The speaker had to name the ink color of the written word. When the ink color was incongruent with the word (e.g., the word red printed in green ink) the speaker should have encountered lexical competition. The deadline to begin speaking was determined by the average speech onset time in a previous experiment using the same task. When the speaker had to beat the deadline and the ink color was incongruent, the speaker postponed part of word planning after the onset of speech. This led to long word durations.

An alternate view comes from Ferreira (2007). She argues that the acoustic characteristics of an utterance should be influenced differently according to their source. She argues that planning should influence only pausing. Word duration and pitch should be only determined by linguistic rules. This view is based on some of her previous findings. Ferreira (1991) manipulated syntactic complexity of subject and object phrases in rehearsed utterances. The following utterances are ordered from least to most syntactically complex. The subject phrase always precedes the object phrase.

(1) The enthusiastic band pleased the very important crowd

(2) The pianist in the band pleased the senator in the crowd

(3) The man who was in the band pleased the girl who was in the crowd

She then measured the duration of the head nouns of each phrase (the underlined words). She found no effect of syntactic complexity on word duration even though she found an effect on pause duration according to object complexity (i.e., the pause between band and pleased was longest in (3) compared to (1)). However, Ferreira (1993) manipulated prosodic phrasing and found that words (e.g. *cop*) at prosodic phrasing boundaries were longer compared to words at boundary internal locations. This is predicted by linguistic rules for phrase final lengthening. Her manipulation is exemplified in the following two utterances:

(1) The friend of the cop infuriated the boyfriend of the girls

(2) The cop who's a friend infuriated the boyfriend of the girls.

In (1) cop is at the boundary of the first prosodic phrase. In (2) cop is boundary internal. The combination of results by Ferreira (1991, 1993) suggests that prosodic rules influence word duration but syntactic complexity does not. Syntactically complex utterances are difficult to plan therefore planning is not related to word duration.

The most important criticism for Ferreira's (2007) conclusion about the role of planning on word duration is that she uses evidence based on prepared speech. Even though this task encourages the access of the full linguistic structure of an utterance, it arguably minimizes the effect of planning. Furthermore, it is possible that word duration variation effects occurred earlier in the phrase and not at the phrase final word.

Further support for the view that planning is unrelated to word duration variation is provided by Damian (2003). Damian (2003) failed to replicate the Kello et al. (2000) study using the same Stroop task, but also using three additional lexical interference tasks. However, upon comparing the deadlines between experiments, the deadline was earlier in the Kello et al. (2000) study. Therefore, the Kello et al. (2000) participants might have been more prone to plan less prior to speech onset, compared to the Damian (2003) participants. The deadline for the Kello et al. (2000) was 589 ms whereas the deadline in Experiment 4 for the Damian (2003) study was 642 ms. Nevertheless, the fact that the effect was not replicated in three variants of the task might suggest that the effect only occurs under very specific circumstances, such as when speakers are forced to start speaking before they are fully prepared.

Further support for the lack of a planning effect on word duration is provided by Schriefers & Teruel (1999) who used a standard picture-word interference (PWI) task. Speakers named the color of a depicted picture along with the depicted object e.g., *red dog*. They analyzed the duration of both words in their first experiment. However, they failed to find any evidence that planning facilitation leads to word duration variation of either word. Nevertheless, their evidence must be treated with caution because the coding of word boundaries was done perceptually and not based on acoustic landmarks.

In sum, the word duration variation and planning literature suggests a relationship between planning and word duration. Planning facilitation should lead to short word duration whereas planning difficulty should lead to long word duration (Arnold & Watson, under review; Balota et al., 1989; Bell et al, 2003, 2009; Clark & Fox Tree, 2002 Christodoulou, 2009; Fowler & Housum, 1987; Fox Tree & Clark, 1997; Kello et al., 2000). However, this proposal is disputed on the grounds that syntactic complexity or planning inhibition may not impact word duration (Damian, 2003; Ferreira 1991, 1993, 2007; Schriefers & Teruel, 1999).

Evidence from the scope and timing of planning literature in conjunction with the word duration variation literature suggests an inverse relationship between planning variation and word duration (Christodoulou, 2009; Griffin, 2003; Ferreira & Swets, 2002; Wagner et al., 2010). However, this inverse relationship might only occur when speakers are forced to either preplan but preplanning is difficult (Griffin, 2003). It might also occur when speakers are forced to plan incrementally and incremental planning is difficult (Ferreira & Swets, 2002). In other words, there might need to be tension between when speakers *should* plan and when they *can* plan. Therefore, the direction of the effect of planning variation on word duration cannot be securely determined. It is possible that in a different task the timing of planning and word duration could covary. This might be the case when the timing of planning occurs early but the speaker cannot easily resolve issues with planning difficulty.

Therefore there are two competing predictions about the relationship between planning variation and word duration variation. The two variables could either have an inverse relationship or they could covary.

The purpose of the following studies is to address the debate about the influence of planning on word duration. Furthermore, they will address the question of whether the timing of planning affects word duration and whether it interacts with the effect of planning on word duration. Lastly, even though there is evidence in the literature that difficulty of an upcoming word leads to word duration variation (Bell et al., 2003; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997), there is no evidence from a well-controlled experimental task.

The purpose of the experiments described in the following chapters is to test two hypotheses: (1) planning a difficult upcoming word leads to long duration for the current word compared to planning an easy upcoming word (2) variation in the timing of planning the upcoming word leads to word duration variation.

Experimental paradigm

In the experiments reported here, I used a picture-naming task previously used to answer questions about incremental planning (Griffin, 2003; Meyer, 1996, 1997; Meyer et al., 1998; Meyer, Roelofs, & Levelt, 2003; Meyer et al., 2007). I presented participants with pairs of objects that consisted of a picture on the left side and a picture on the right side of a computer screen. The picture on the left always had a low frequency name, whereas the picture on the right either had a low or a high frequency name. The participant's task was to name the objects from left to right while trying not to pause (Griffin, 2003; Meyer et al., 2007).

The advantage of using the dual-picture naming task instead of analyzing naturally occurring speech is that I can manipulate specific characteristics of the pictures and measure the timing of planning using the eye-voice span (e.g., Grfiffin & Bock, 2000). For the purposes of this dissertation, the eye-voice span will be defined as the difference between the onset of the first fixation to the critical object (i.e. the right object in the display) and the onset of speech. The definition of this measure is different that Griffin & Bock (2000) who defined it as the difference between that onset of the last fixation to the critical object and the onset of the noun referring to the target picture. Furthermore, by keeping the utterance to only two words I minimize any possible confounds with utterance complexity measures such as utterance length. Additionally, I am able to minimize any influence of a prosodic representation on word duration variation, such as pre-boundary lengthening (e.g., Ferreira, 1993).

Lexical planning requires longer processing time when word frequency is low as opposed to when it is high (Oldfield & Wingfield, 1965). Therefore, in the low frequency condition the speaker should require more processing time compared to the high frequency condition either in the form of pausing before Word1 (the name of the left object) or in the form of extending Word1 duration to plan Word2 (the name of the right object). Therefore, planning low frequency words should take longer compared to planning high frequency words.

According to theories of incremental planning, early planning stages of the right word can overlap with the articulation of the left word (e.g., Kempen & Hoenkamp, 1987; Meyer & Levelt, 2000). The production system can thus minimize working memory requirements of the task by distributing lexical planning over an extended period of time (Levelt & Meyer, 2000). For example, incremental planning allows working memory to free up resources to plan the right word by pushing the left word into the articulatory process. Nevertheless, second word planning cannot be completely postponed after the first word is fully articulated. This would require a significant pause between the two words that would violate task constraints.

In the current task, the speaker will then have the choice to either preplan both words or to plan part of Word2 after speech onset. If the speaker preplans both words, there should be little need for incremental planning. Therefore, the need for word duration variation should be minimized. If the speaker plans part of Word2 after speech onset, the need for incremental planning should be higher.

A difference in speech onset time according to Word2 planning difficulty will constitute evidence for preplanning. The absence of a by-condition difference on speech onset time will constitute evidence for incremental planning.

However, even if speakers have a dominant response of either preplanning or incremental planning, there could still be some variation on the timing of Word2 planning. Therefore, independent of the dominant type of planning, word duration could fluctuate. For example, according to one outcome speech onset time might not vary according to Word2 planning difficulty. Even if there is evidence for incremental planning, the speaker on certain trials might plan less incrementally. This could shift the timing of Word2 planning earlier than the usual timing determined by incremental planning.

To test my main hypotheses first I need to identify how speakers perform in my version of the dual-picture naming task. The purpose of Experiment 1 is to address this issue. Previous literature suggests that Word2 planning does not affect speech onset time (e.g., Meyer et al., 2007). Therefore speech onset time should not differ according to Word2 frequency in the current task. However, this effect is arguably mediated by the degree of pausing and Word1 duration variation. These factors were not statistically accounted for in previous studies. It is possible that speech onset time will vary once other confounds are accounted for.

Previous literature also suggests that planning difficulty affects word duration (Bell et al., 2003; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). Word1 duration should be long when Word2 is low frequency as opposed to high frequency. However, this effect has not been previously tested in an experimentally controlled task. It is possible that word duration variation does not occur when speakers plan a sequence of unrelated bare nouns. This will therefore be the first test of the first hypothesis.

The purpose of experiments 2-4 is to accurately measure the timing of Word2 planning and to impose additional manipulations on the timing of planning. Experiment2 is identical to experiment1 with the addition of the eye-voice span (e.g., Griffin & Bock, 2000). In Experiment 3 I manipulate the time available for incremental planning by allowing speakers to introduce the conjunction *and* between the two bare nouns. This manipulation is informed by the second experiment in Griffin (2003). She showed that the relationship between Word1 duration and the timing of planning disappears when speakers can add the phrase *next to* between the two bare nouns. In Experiment 4 I manipulate the time available for preplanning by imposing two different deadlines to begin speaking. This manipulation is informed by the deadline procedures used in the literature that suggests speakers use long word duration when the deadline is short (Ferreira & Swets, 2002; Kello et al. 2000).

CHAPTER 2

EXPERIMENT 1: THE EFFECT OF WORD2 FREQUENCY ON WORD1 DURATION

The purpose of Experiment 1 was to establish how speakers respond in the current version of the dual-picture naming task. It also served the purpose of testing the first hypothesis that difficulty of planning Word2 will lead to Word1 duration variation.

Speakers named two pictures from left to right while trying not to pause. The name of the picture on the right varied in frequency (high vs. low) as a means of varying planning difficulty. Low frequency should lead to a higher level of planning difficulty compared to high frequency. I predicted that the low frequency condition would lead to long Word1 duration compared to the high frequency condition. I measured speech onset time, word duration, and pause duration.

There are two possible outcomes that could fail to support the role of planning on word duration. One possible outcome (outcome 1) is that neither speech onset time nor Word1 duration would vary according to Word2 planning difficulty. A second possible outcome (outcome 2) is that speech onset time would vary because of Word2 planning difficulty but Word1 duration would not vary.

There are also two possible outcomes that could support the role of planning on word duration. One possible outcome (outcome 3) is that speech onset time would not vary but Word1 duration would vary because of Word2 planning. A second possible outcome (outcome 4) is that both speech onset time and Word1 duration would vary.

Both outcomes 3 & 4 could provide evidence for the mechanism that leads to word duration variation. Evidence for outcome 3 would suggest the speaker was most probably planning incrementally, because Word2 did not influence speech onset time. Therefore, word duration variation could potentially occur under circumstances where incremental planning is taxed. In other words, Word2 duration variation could occur when the speaker plans one word at a time but then needs extra time during Word1 articulation to plan Word2.

Evidence for outcome 4 would suggest the speaker was both preplanning and planning incrementally. Therefore according to this account word duration variation could occur when speakers are potentially slowing down their overall rate of processing. Furthermore, according to this outcome the speaker could have some information about Word2 prior to speech onset.

Method

Participants

A total of 27 students from the community at UNC Chapel Hill participated in the experiment in exchange for course credit. Three participants were excluded from the analysis and replaced because of equipment failure. All participants were native speakers of English or bilinguals of English and another language.

Equipment

I used the experimental software Psycript 2.3.0 (Slavin, 2006) installed on a Mac mini (OSX 10. 4.11). The stimuli were presented on an IBM 23 in. monitor. The participants' spoken utterances were recorded using an Audio-Technica head-worn cardioid condenser microphone, attached to a Marantz portable solid-state recorder. The recorded utterances were auto-segmented with the P2FA set of scripts (Yuan & Liberman, 2008) and analyzed after checking for segmentation errors using the software program Praat (Boersma & Weenink, 2011). Data management and analysis in Praat was aided in part by the GSU Praat tools Owren (2008) and in part by Praat scripts developed by the author.

Materials

I used colored object drawings taken from the Rossion & Pourtois (2004) pictorial set. The Rossion & Pourtois (2004) set is a colorized version of the non-colorized line drawings created by Snodgrass & Vanderwart (1980). I chose the colorized picture set as opposed to the non-colorized set because according to Rossion & Pourtois (2004) color improves perceptual recognition and naming agreement of the non-colorized objects of Snodgrass & Vanderwart (1980).

I used a total of 72 drawings, which I divided into three sets. The first set comprised 24 pictures with low frequency names to be presented on the left side of the screen. I will refer to this as the L-target set. I used pictures with low frequency names as opposed to high frequency names to maximize the difficulty of full utterance preplanning.

The second and third set each comprised 24 pictures to be presented on the right side of the screen. The second set comprised pictures with low frequency names and the third set comprised pictures with high frequency names. I will refer to set two as the low frequency set and set three as the high frequency set. These two sets defined the two experimental conditions (high vs. low frequency names).

Items were assigned to each frequency group on the basis of frequency counts reported in Snodgrass & Vanderwart (1980). I selected 72 items with equal numbers of monosyllabic and disyllabic words. I then selected 36 items with the highest log frequency values for each syllable length for the high frequency group (range: 1.36 - 2.88). I selected 36 items with the lowest frequency values for the low frequency groups (range: 0 - 1.11). I tested whether the high and low frequency right object sets differed in frequency with a one-way ANOVA (three levels: L-target set vs. low frequency vs. high frequency) (F (69,2) = 150.28, $p < 0.001$). Post hoc comparisons using the Tukey HSD test indicated that the mean log frequency of the L-target set (M = 0.35, SD = 0.39) and the low frequency set (M = 2.14, SD = 0.44) were significantly different from the high frequency set (M = 0.38, SD = 0.39) but did not significantly differ from each other.

The three sets were matched on other measures that could influence the acoustics of the spoken utterances such as name agreement, visual complexity, and concept-image agreement. Name agreement quantifies difficulty in uniquely identifying a name for an object. I used the H statistic of name agreement (Snodgrass & Vanderwart, 1980), which is the percentage of the most frequent name adjusted for the number of different names provided for the same object. High name agreement (e.g., 0) is indicative of a high chance in uniquely identifying a name. Visual complexity quantifies difficulty in processing an object perceptually. This was measured by asking participants to report on a 5-point scale (1 simple, 5 complex) their perceived complexity of the depicted object (Snodgrass & Vanderwart, 1980). Typically, complex objects require more visual processing time compared to less complex objects. Concept-image agreement quantifies the difficulty in recognizing an object given prior expectations about what the object should look like. This was measured by asking participants to visualize a concept and then rate the match on a 5-point scale (1 low agreement, 5 high agreement) between the visualized object and the object they were presented with (Snodgrass & Vanderwart, 1980). All ratings were obtained from Rossion & Pourtois (2004) except for the H statistic, which was obtained from Snodgrass and Vanderwart (1980). Means and standard deviations for picture characteristics according to set are provided in table 1.

Table 1

Means (and standard deviations) for properties of critical objects

	Left Object		Right High Frequency		Right Low Frequency	
	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Log word frequency (SV80)	0.35	0.39	2.14*	0.44	0.38	0.39
Codability (SV80)	0.43	0.42	0.41	0.44	0.36	0.36
Codability (RP04)	0.43	0.51	0.54	0.79	0.28	0.41
Visual Complexity (RP04)	2.84	1.05	2.48	0.77	2.91	0.94
Imageability (RP04)	3.90	0.58	3.56	0.80	3.84	0.87
Familiarity (RP04)	3.03	1.09	4.33*	0.75	2.92	0.91

Note. SV80 refers to picture characteristics as reported in Sondgrass and Vanderwart (1980). RP04 refers to picture characteristics as reported in Rossion and Pourtois (2004). Mean Values with an asterisk are significantly different from all other mean values in the same row.
*p<0.01

In sum, the materials were controlled for the following characteristics: difficulty of retrieving the appropriate name of an object, difficulty in visually identifying the object, surprise in the way the object is depicted. The frequency manipulation was confounded with conceptual familiarity, which is to be expected. This was confirmed with a one-way ANOVA (F (2,69) = 17.16, $p < 0.001$). A post hoc Tukey HSD test indicated that the familiarity of the L-target set ($M = 3.03$, $SD = 1.09$) and the low frequency set ($M = 2.92$, $SD = 0.91$) was significantly different than the high frequency set ($M = 4.33$, $SD = 0.75$) but the two sets did not differ from each other. This difference does not invalidate our frequency manipulation because its purpose is to modulate preplanning difficulty. However, it does suggest that the locus of any outcome is not restricted to one level of planning. Frequency effects are typically attributed to the phonological planning stage (Jescheniak & Levelt, 1994; Jurafsky, Bell, & Girand, 2000) and possibly at the lexical retrieval stage (Alario, Costa, & Caramazza, 2002; Dell, 1990; Gahl, 2008). The difference in conceptual familiarity in addition to name frequency suggests that the locus of any possible outcome could potentially originate before the lexical retrieval stage.

Design

I used the three picture sets to create pairs of pictures. One picture from the L-target set was paired with one picture from the low frequency and one picture from the high frequency set while keeping the number of syllables constant. I matched the paired pictures according to number of syllables, similar to Meyer et al.'s (2007) second experiment. Meyer et al. (2007) manipulated the age of acquisition (AoA) of the objects on the left side of the screen, which is highly correlated with frequency. All target words of objects on the left and right side were disyllabic except for two. The purpose of matching all words on number of syllables was to isolate the effect of planning difficulty

(AoA) of the left picture name on speech onset. Their experimental results showed that the speech onset time in the early AoA condition was shorter by 27 ms compared to the late AoA condition. Their results were significant by subjects and marginally significant by items.

This procedure resulted in two matched picture pairs, which I treated as an item. For example, the picture of a skunk from the first set was paired with the picture of a deer from the second set and the picture of a hand from the third set. These matched drawing pairs were treated as a single item as depicted in figure 3. This design afforded a within item comparison between high and low frequency words. Overall there were 24 items that occurred in both frequency conditions (48 distinct pairings).

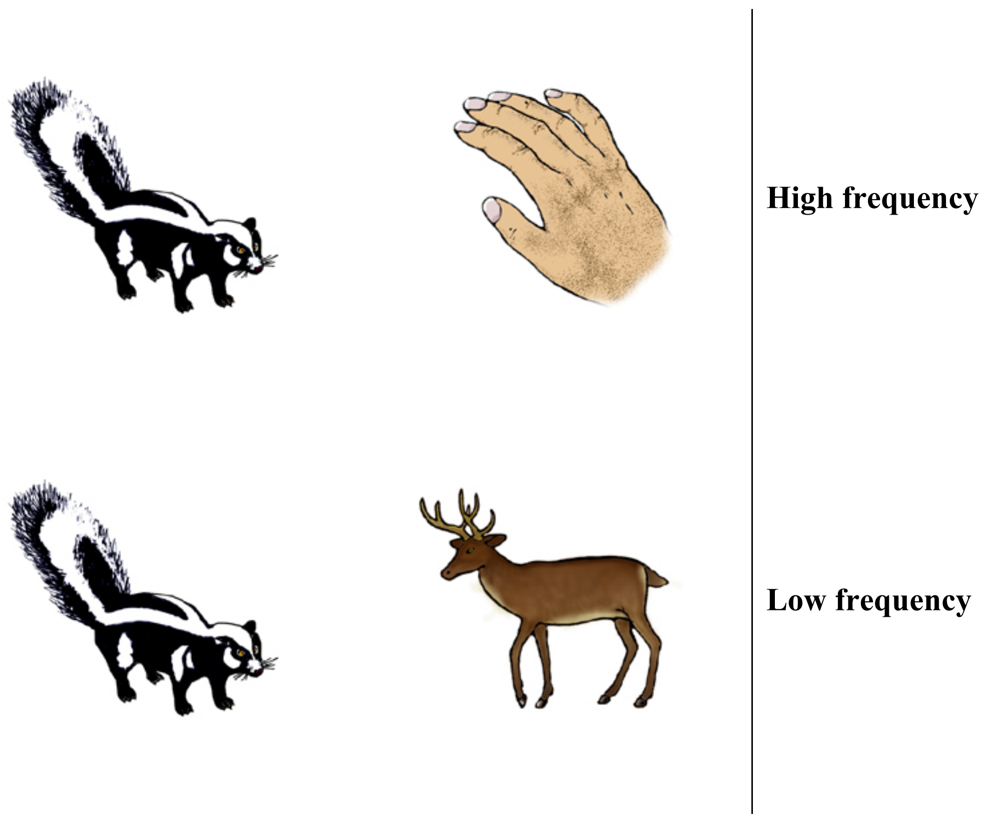


Figure 3. Example of an experimental item. This figure depicts a single item that occurred in both frequency conditions. The picture on the left (skunk) was paired with two other pictures (hand, deer).

I then created two lists where each column represented a frequency condition and each row an item. I combined the two lists by using a Latin Square. This methodology allows the creation of two experimental lists where items are not repeated but the participant is exposed to equal numbers of each level of the independent variable. Therefore, the participant is exposed to all items but only in a single condition.

I also tried to encourage a trial-by-trial evaluation of difficulty by inserting an additional 22 filler items. The filler items were created by combining two picture sets with varying number of syllables. Thus, the resulting pairs varied in syllable length. Half of the pairs always had a multisyllabic left name and a short right name (monosyllabic or disyllabic). The other half of the pairs had always a short left name (monosyllabic or disyllabic) and a multisyllabic right name. All fillers were present in both experimental lists.

Procedure

Participants were tested individually. They were seated in front of a computer screen at a distance of approximately 2 feet where they could clearly see the stimulus pictures. First, a lab assistant or I gave them a brief overview of the task followed by instructions about performing the task. We explained that they had to try not to pause between the two words and that they would have a chance to practice the task for 12 trials. During the practice trials we gave participants feedback if they were pausing significantly between the two words or if they were inserting extra words such as *uh, um, and, the*. None of the practice items appeared in the experimental session. After completing the practice session we informed the participants that the actual experiment would take place and that it would take longer than the practice session. I rotated the experimental lists for each participant. Furthermore, the experimental software randomized the order of trial presentation.

For each trial the sequence of events was the following. The participants first saw a red X in the middle of the screen. They were instructed to click on the X to initiate the trial. They would then immediately hear a beep and the two pictures appeared on the left and right part of the screen. The purpose of the beep was to inform them that the trial had started, however they were not told they needed to start speaking as soon as the beep sounded. They then named the two pictures and clicked on either of the two, to move onto the next trial. This sequence of events was identical for all trials.

When the session was over, we asked the participants a series of questions about the experiment and gave them a full debriefing on the purpose of the experiment. The recordings were then transcribed and submitted to acoustic analysis.

Statistical Analysis

I analyzed the data using mixed effects models with subjects and items treated as cross-classified random variables (Baayen, Davidson, & Bates, 2008). These are general linear models that can address issues of clustering of observations within datasets. The clustering is a result of the nesting of observations within speakers and within words. For example, speaker Miri can have a general tendency to use long words. However, speaker Vicki can have a general tendency to use short words. These characteristics will result in positive intra speaker word duration correlations but negative inter speaker word duration correlations.

I built separate models for each outcome variable (Word1 duration, Speech onset time). The models included two different types of parameters, random effects and fixed effects. The fixed effects were further divided into control variables and predictor variables that were theoretically motivated. In my design I had four different sets of control variables. The first set included variables that were the result of the experimental design (block, list, trial order, number of syllables, initial and final phoneme category). The second set included variables coding for the initial and final phoneme category of Word1 and Word2. I used the following categories: 1. Vowel, 2. Stop, 3. Affricate, 4. Fricative, 5. Liquid, 6. Nasal. The third set included participant characteristics such as sex and outcome variables that were the result of performing the required task (subject error rate, item error rate, subject disfluency rate, item disfluency rate). The third set included picture characteristic variables (codability, familiarity, imageability, visual complexity).

The random effects included subject and item id variables that introduced individual intercepts for each subject and item. I also included slope adjustments to each intercept where it was warranted by the design. The individual slopes represented differences in the response of each subject and item to the independent variable(s). For example speaker Miri who uses overall long words, might show a larger difference in word duration between the two frequency conditions compared to speaker Vicki who uses overall short words. The prerequisite for including a random slope was for each speaker to have at least one observation in each design cell. This was not feasible in cases where the experimental design did not allow this or in cases where a subject or an item was missing all data in one of the cells.

The model fitting process included four different phases. First, I introduced the random effects structure. I then introduced the control variables according to set number order. Thirdly, I introduced the predictor variables. At the last step I conducted model criticisms to inspect for outliers that might be influencing the fit of the model.

The model fitting process was guided by considerations for parsimony and for colinearity. According to parsimony, the model with the least number of parameters justified by the complexity of the data should be retained. Colinearity refers to the correlation between predictor variables. The result of colinearity is the explanation of non-unique variance in the outcome variable, which tends to inflate type 2 error. To mitigate any issues of colinearity I centered the control and predictor variables. Furthermore, I only retained control variables that were not collinear with any predictor variables. However, control variables were allowed to be collinear with each other.

At each phase I created successively nested models. This means that each superordinate model was identical to the following subordinate model except for a minor change in parameter specification. I then retained the most parsimonious model (the model with the least number of parameters) if a more complex model was not explaining a significantly higher amount of variance. This criterion was implemented in two different ways. When introducing fixed effects I only retained parameters associated with a t value of 1.5 or higher. When introducing the random effects structure I conducted successive log likelihood tests for each nested model. This allowed me to eliminate correlations between intercepts and slopes. Furthermore, I only retained control variables that had a partial correlation value with a predictor variable of less than 0.2. This last criterion addressed any issues of colinearity between the fixed effects.

None of the theoretically important variables had a correlation higher than 0.2. Only interaction terms between two predictor variables and a predictor variable in the interaction, or two interaction terms with shared predictors were allowed to have a partial correlation higher than 0.2. None of those correlations were higher than 0.5. Serious colinearity issues arise when predictor variables have a partial correlation of 0.7 or higher and only result in loss of power.

Results

The experimental results show that the low frequency condition led to longer Word1 duration than the high frequency condition. However, Word2 frequency did not affect speech onset time. This pattern of results is depicted in Figure 4. All significant control variables for both Word1 duration and speech onset time are listed in table 2.

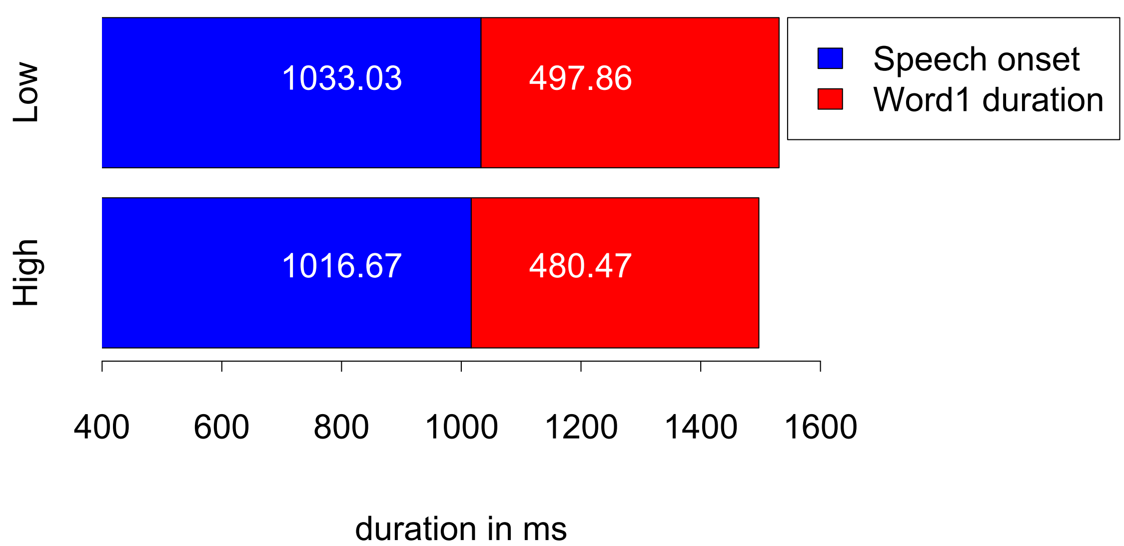


Figure 4. Experiment 1 Speech onset time and Word1 duration sample based means according to condition

Table 2

Control predictors for each model in Experiment 1

Variable types	Predictors	Word1 duration	Speech onset
Design	Number of syllables	X	
	List		
	Trial	X	X
	Pause	X	X
Phoneme	Word1 onset	X	
	Word1 offset		
	Word1 offset*Word2 onset		
Performance	Sex		
	Error rate by subject		
	Error rate by item		
	Disfluency rate by subject		
	Disfluency rate by item		
Characteristics	Frequency		
	Word 1 Familiarity		
	Imageability		
	Word 2 Codability		X
	Word1 Visual complexity		

Note. Design variables refer to variables relevant to the design of the experiment. Phoneme variables refer to variables relevant to the onset and offset of Word1 and Word2. Performance variables refer to speaker characteristics and error and disfluency rate. Characteristics refer to Word1 or Word2 picture characteristics.

Exclusions and error rate analysis

The overall error rate before exclusions was (18% or 106 trials out of 576 trials). Items and subjects with a combined naming and accuracy error rate higher than 33% were excluded from further analyses. This threshold of 33% has been previously used in experiments with high memory load (e.g., Wagner et al. 2010). This led to the exclusion of three items (72 trials or 12.5% of the original data) and the replacement of two participants. After the exclusions the error rate was reduced to 12.5% (or 72 trials). I also excluded trials with disfluencies. I considered a disfluency the insertion of a word between the target words or before the first target word, such as *um, uh, and, a, the* or any mispronunciation of either of the two target words. This led to the exclusion of 24 trials (4.2% of the original data). Furthermore, I excluded trials with speech onset time longer than 2000 ms (20 trials or 3.5% of the data) and pause duration longer than 200 ms (52 trials or 9% of the data). Lastly I excluded four additional observations because of miscellaneous errors. Overall I excluded 29% of the data (or 168 trials out of 504) after excluding the three items or 41.6% (or 240 trials out of 576) including the three excluded items.

I compared the error rate across frequency conditions using F1 and F2 ANOVAS. The error rate did not differ between the low (M =10%, SD = 5%) and high (M = 10%, SD = 6%) frequency conditions ($F1$ (1,51) = 1.13, p = 0.29, $F2$ (1,40) = 0.33, p = 0.57). Also, disfluency rate did not differ between the low (M = 2%, SD = 3%) and the high (M = 2%, SD = 3%) frequency conditions ($F1$ (1,51) = 0.01, p = 0.91, $F2$ (1,40) = 0.29, p = 0.59).

Word1 duration analysis

Word2 frequency influenced Word 1 duration. The final model for log word duration included random intercepts for items and subjects with uncorrelated random slopes for condition. The model included controls for number of syllables, trial order, duration of pausing, phoneme category of Word1 onset.

The predictor was Word2 frequency category, which had a significant effect on the log of Word1 (β = 0.0466, t = 4.43, p < 0.001). To interpret model-implied effects I used the effects package (Fox, J., 2003). When computing the predicted values for a specific predictor variable, the effects package holds all other control variables at their median. I then back-transformed the data to the original scale. Word1 duration was

49.32 ms longer when Word2 was low frequency as opposed to high frequency. All fixed effects and their associated significance values can be found in table 3.

Table 3

Experiment1 Word1 duration fixed effect model estimates and significance values after MCMC sampling

	b	*SE*	*t*	*p*
Intercept	2.6602	0.0169	157.5500	0.0000
Syllables	0.0673	0.0245	2.7400	0.0064
Pause	0.0113	0.0061	1.8400	0.0669
Trial#	0.0004	0.0003	1.2200	0.2227
W1onset	0.0247	0.0103	2.4000	0.0170
W2freq	0.0466	0.0105	4.4300	0.0000

Note. Trial# refers to the order of trial presentation. W1onset refers to the phoneme category of Word1. W2freq refers to W2 frequency category

Speech onset time analysis

Word2 frequency did not influence speech onset time. The final model for log speech onset time duration included random intercepts for subjects and items with uncorrelated random slopes for condition, only for item intercepts. I did not include random slopes by condition for subjects because they were accounting for 0 variance. The model included controls for duration of pausing, trial number, and codability for Word2. The predictor was Word2 frequency category, which did not have a significant effect on the log of speech onset time ($\beta = 0.0133$, $t = 1.18$, $p = 0.24$). All fixed effects and their associated significance values can be found in table 4.

Table 4

Experiment 1 Speech onset time fixed effect model estimates and significance values after MCMC sampling

	b	*SE*	*t*	*p*
Intercept	3.0124	0.0174	172.7600	0.0000
Pause	-0.0114	0.0056	-2.0400	0.0422
Trial#	0.0009	0.0003	2.8500	0.0047
W1H	0.0722	0.0212	3.4000	0.0007
W2frequency	0.0133	0.0113	1.1800	0.2376

Note. Trial# refers to the order of trial presentation. W1H refers to codability of Word1.

Discussion

The results of Experiment 1 provide evidence that speakers were planning incrementally. Word1 duration was long when followed by a low frequency as opposed to a high frequency word. However speech onset time did not differ. The finding that speakers were planning incrementally is consistent with previous evidence in the literature that the dual-picture naming task encourages incremental planning as opposed to complete utterance preplanning (e.g., Meyer, 1996). The results are compatible with outcome 3 that supports the role of planning on word duration.

Arguably, planning incrementally is rather risky in this task because the speaker bets that he or she can plan at least most of the phonological information of Word2 without a complete estimate of how much time Word2 planning will take. Therefore, the speaker runs the risk of failing to meet the task demands of not pausing between the target words. This strategy could be potentially motivated by the benefit of incremental planning of minimizing Word1 buffering (e.g., Griffin, 2003; Levelt & Meyer, 2000). If we assume that speakers are planning serially, the speaker will have to buffer Word1 while retrieving information about Word2. Incremental planning would then be preferred if speakers favor minimizing word buffering over ensuring fluent delivery.

Another reason why incremental planning could be preferred is possibly because the limited look-ahead of Word2 planning is adequate to accommodate fluent delivery. In a secondary analysis on Word1 duration I also included observations with pauses longer than 200ms. I then introduced an interaction term between Word2 frequency and the presence of a pause longer than 200 ms.

The effect of Word2 frequency on Word1 duration was robust even after including trials with pauses longer than 200 ms and independent of the presence of those pauses. The fixed effects of the model can be inspected in Appendix B. These results suggest that even in cases where words were not closely planned together, the speaker at least began planning Word2 during the articulation of Word1. Therefore, the limited look-ahead could accommodate planning coordination even in cases that could be categorized as disfluent.

In sum, in the first experiment speakers planned incrementally. The first experiment provided evidence that Word2 planning difficulty affected Word1 duration. Low Word2 frequency led to long Word1 duration.

CHAPTER 3

EXPERIMENT 2: THE EFFECT OF WORD2 FREQUENCY AND REVS ON WORD1 DURATION

The purpose of Experiment 2 was to test both hypotheses that Word2 planning difficulty and the timing of Word2 planning affects Word1 duration. In Experiment 2 I used the same dual-naming task as Experiment 1 and I also added an eyetracking measure to identify the precise timing of Word2 planning.

Experiment 1 results suggested that speakers only planned Word1 prior to speech onset. However, this does exclude the possibility of trial-by-trial variation in how much Word2 information was available prior to speech onset. In other words, even though on average speakers planned only Word1 prior to speech onset, on some trials speakers could have planned some information of Word2 prior to speech onset.

As an index of the point when Word2 planning began, I measured the right object eye-voice span (REVS), which is the difference between the time of the first fixation to the right object with relation to the onset of speech (e.g., Griffin & Bock, 2000). A negative REVS (Early REVS) suggests that the timing of Word2 planning occurred before the onset of speech. A positive REVS (Late REVS) suggests that the timing of Word2 planning occurred after the onset of speech. In figure 5 I depict only two scenarios for REVS as either early or late for the sake of clarity of representation.

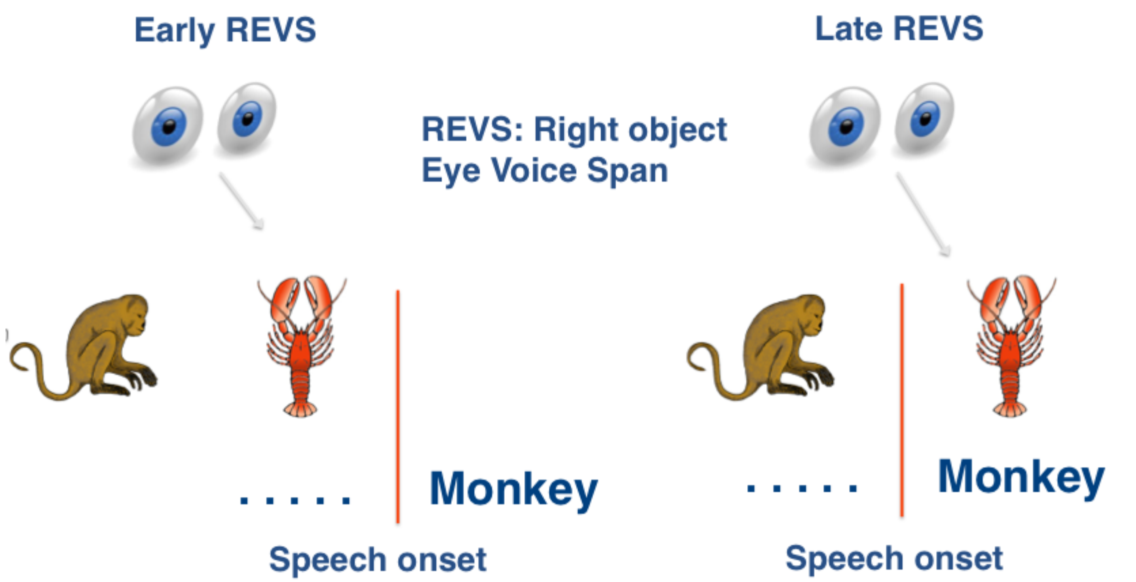


Figure 5. Depiction of two possible REVS Scenarios. Early REVS occurs when the speaker looks at the right object (the lobster) before speech onset. Late REVS occurs when the speaker looks at the right object after speech onset

However, the actual measure is continuous. Therefore, the measure captures variation in early and late REVS.

Based on Experiment 1 results I predicted long Word1 duration when Word2 is low frequency as opposed to high frequency. Speech onset time should be unrelated to Word2 frequency.

I also tested the hypothesis that the timing of Word2 planning would lead to Word1 duration variation. One possible scenario of how the timing of Word2 planning could relate to Word1 duration is that early timing of planning would lead to Word1 reduction. If speakers have some information about Word2 in a subset of trials prior to speech onset, this should reduce the need for incremental planning. Furthermore, the combination of the variation in planning literature and word duration variation literature suggests an inverse relationship between the timing of planning and word duration (Christodoulou, 2009; Griffin, 2003; Ferreira & Swets, 2002; Wagner et al., 2010).

A different possible pattern of results is that of covariation between REVS and Word1 duration. Early REVS could lead to long Word1 duration. When information about Word2 is available prior to speech onset, Word1 is long. This pattern would suggest that during certain trials, early timing of planning is also associated with extensive post speech onset planning. For example this scenario could occur if speakers are both preplanning and planning incrementally certain words, potentially because they are encountering high cognitive load.

Method

Participants

A total of 33 students from the community at UNC Chapel Hill participated in the experiment in exchange for course credit. 10 participants were excluded from the analysis and replaced because of equipment failure ($n = 8$), for high naming error ($n = 2$), for not being a native speaker of English *($n = 1$).* All other participants were native speakers of English or bilinguals of English and another language. The equipment failure rate was mainly due to a faulty microphone issue ($n = 7$). The audio recordings were contaminated with artifact noise that masked the acoustics of the target words. The remaining equipment failure was due to the lack of eyetracking data output *($n = 1$).*

Materials and Procedure

Experiment 2 was identical to experiment 1 except for the addition of eyetracking. Participants wore an Eyelink II eyetracker attached to two PCs. The eyetracker was calibrated using a Dell PC running proprietary Eyelink software while the experimental stimuli were presented on an IBM PC running Windows XP. The stimuli were presented on a 19’’ screen with 1280X765 resolution with the experimental software Exbuilder (Longhurst, 2006).

All other equipment, material and procedures were identical to Experiment 1 except for the following changes. I replaced four stimulus pictures that were associated with a high error rate in Experiment 1. A list of all the changes in stimuli across all experiments is provided in the appendix A. These changes did not lead to any changes in the picture characteristic differences between conditions. Frequency and conceptual familiarity were significantly different between the low frequency picture sets and the high frequency picture set.

The procedure was different compared to experiment 1 in the following respects. Before each trial the participants fixated on a dot in the middle of the screen. During the fixation the Eyelink software conducted a drift correction that adjusted for any minor changes in the position of the eyetracker with relation to the screen. The trial did not begin until the software had detected a stable fixation on the dot to conduct the drift correction. The duration of this event varied according to the quality of the track. Participants then completed the task by naming the two pictured objects from left to right without pausing as in Experiment 1. I used the data analysis software Dataviewer by Eyelink to analyze the fixation data.

Results

The experimental results show that low Word2 frequency led to longer Word1 duration than high Word2 frequency. However, Word2 frequency did not affect speech onset time. Both of these results are consistent with Experiment 1 results. The pattern of results is depicted in Figure 6. The timing of Word2 planning significantly influenced Word1 duration independent of Word2 frequency (see figure 7). All significant control variables for both Word1 duration and speech onset time are listed in table 5.

Exclusions and error rate analysis

The overall error rate before exclusions was 16.32% or 94 trials out of 576 trials. Items and participants with a combined naming and accuracy error rate higher than 33% were excluded from further analyses. This led to the exclusion of two items (48 trials or 8.3% of the original data) and the replacement of two participants. After the exclusions the error rate was reduced to 10.42% (or 60 trials). I also excluded trials with disfluencies. This led to the exclusion of 8 trials (1.9% of the original data). Furthermore, I excluded trials with speech onset time longer than 200ms (20 trials or 3.5% of the data) and with pause duration longer than 200 ms (53 trials or 9.2% of the data). Lastly I excluded four additional trials because of miscellaneous errors. Overall I excluded 26% of the data (or 149 out of 528 trial) after excluding the three items or 34% (or 197 out of 576 trial) of the data when including the three excluded items.

I compared the error rate across frequency conditions. The error rate did not differ between the low (M = 9%, SD = 6%) and high (M = 11%, SD = 8%) frequency conditions ($F1$ (1,44) = 0.53, p = 0.47, $F2$ (1,38) = 0.24, p = 0.62). Disfluency rate did not differ between the low (M = 2%, SD = 3%) and high frequency (M = 1%, SD = 2%) conditions ($F1$ (1,44) = 1.9, p = 0.17, $F2$ (1,38) = 1.66, p = 0.2).

Eyetracking analysis

I used the software Dataviewer by Eyelink to extract the time of the first fixation to the right object. I then subtracted the speech onset time from the fixation time. This is

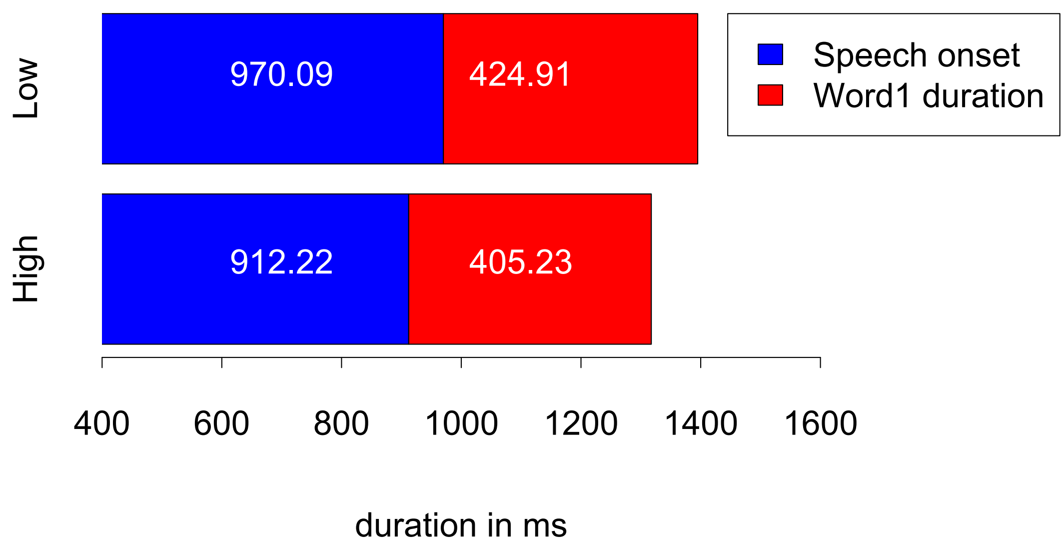


Figure 6. Experiment 2 Speech onset time and Word1 duration sample based means according to condition

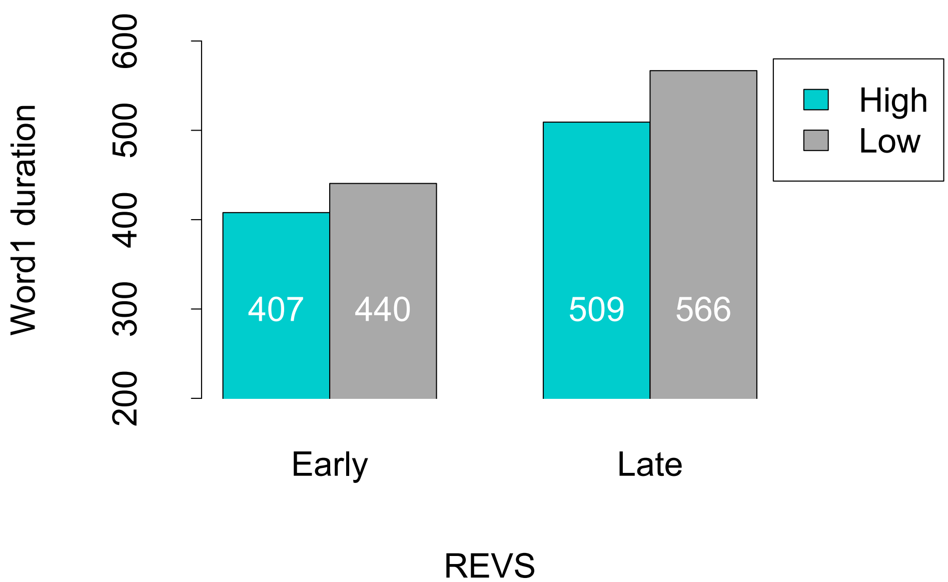


Figure 7. Experiment 2 Word1 duration according to Word2 frequency and REVS. Early REVS consists of looks to the right object prior to speech onset

Table 5

Control predictors for each model in Experiment 2

Variable types	Predictors	Word1 duration	Speech onset
Design	Number of syllables	X	
	List		
	Trial	X	X
	Pause	X	
Phoneme	Word1 onset	X	
	Word1 offset		
	Word1 offset*Word2 onset		
Performance	Sex		
	Error rate by subject		X
	Error rate by item		
	Disfluency rate by subject		
	Disfluency rate by item	X	
Characteristics	Frequency		
	Word 1 Familiarity		
	Imageability		
	Word1 Codability		X
	Word1 Visual complexity		

Note. Design variables refer to variables relevant to the design of the experiment. Phoneme variables refer to variables relevant to the onset and offset of Word1 and Word2. Performance variables refer to error and disfluency rate. Characteristics refer Word1 or Word2 picture characteristics.

the right object eye-voice span (REVS). A positive number indicated that the fixation occurred after speech onset. A negative number indicated that the fixation occurred before speech onset. Speakers showed a preference of looking at the right object after the onset of speech. On average they looked at the right object before the onset of speech 84% of the time in the high frequency condition and 92% of the time in the low frequency condition. I then used this measure to predict Word1 duration and speech onset time.

Word1 duration analysis

REVS and Word2 frequency influenced word duration. The final model for log word duration included random intercepts for items and subjects with uncorrelated random slopes for condition. The model included controls for pausing, number of syllables, trial number, phoneme category of Word1 onset, and disfluency rate by item. The predictors were the REVS and Word2 frequency category. Word2 frequency category had a significant effect on the log of Word1 ($\beta = 0.0255$, $t = 2.51$, $p < 0.05$). To interpret model-implied effects I back-transformed the data to the original scale. Word1 duration was 21.14 ms longer when Word2 was low frequency as opposed to high frequency. REVS[1] also had a significant effect on the log of Word1 duration ($\beta = -0.00007$, $t = 3.48$, $p < 0.001$). Word1 duration was 131.49 ms shorter when REVS was at its maximum negative value as opposed to its maximum positive value. In other words, the earlier looks to the right object with relation to speech onset the more reduced Word1 duration. Fixed effect coefficients and their associated significance values can be inspected in table 6. Model-implied random slopes are depicted in figure 8 and figure 9. The purpose of including both figures is to depict important graphical parameters that I couldn't incorporate into a single figure. In figure 8 I used the effects package (Fox, J., 2003), which provides a rug plot. The rug plot is located just above the X axis and provides information about the relative distribution of REVS. In figure 8 I used the plotLMER.fnc function included in the Language R package (Baayen, 2011). This figure provides confidence intervals in the form of the dashed curves around the model-implied lines.

[1] The effect of REVS also predicts duration when REVS is treated as a categorical variable of looks before vs. after the onset of speech.

Table 6

Experiment 2 Word1 duration fixed effect model

estimates and significance values after MCMC sampling

	b	*SE*	*t*	*p*
Intercept	2.61500	0.02284	114.50000	0.0000
Pause	0.01093	0.00630	1.74000	0.0835
Syllables	0.05346	0.02822	1.89000	0.0592
Trial#	0.00114	0.00030	3.74000	0.0002
W1onset	0.02045	0.01246	1.64000	0.1018
Item dis	0.44810	0.17220	2.60000	0.0097
W2freq	0.02548	0.01016	2.51000	0.0127
REVS	0.00007	0.00002	3.48000	0.0006
W2freq*REVS	-0.00001	0.00003	-0.29000	0.7693

Note. Trial# refers to the order of trial presentation. W1onset refers to initial Word1 phoneme category of. Item dis refers to disfluency rate by item.

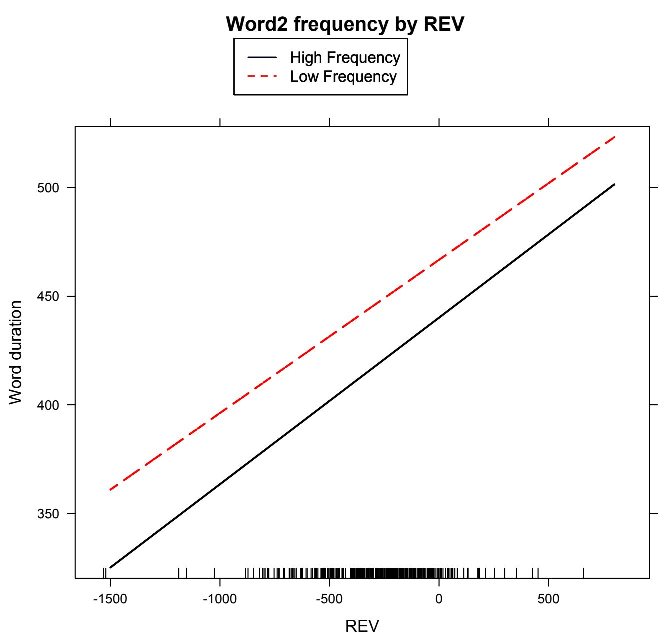


Figure 8. Word2 frequency and REVS predict Word1 duration in Experiment 2. The black solid line represents the model-implied effect of low frequency and the red dashed line represents the model-implied effect of high frequency. The rug plot depicts the relative distribution of REVS

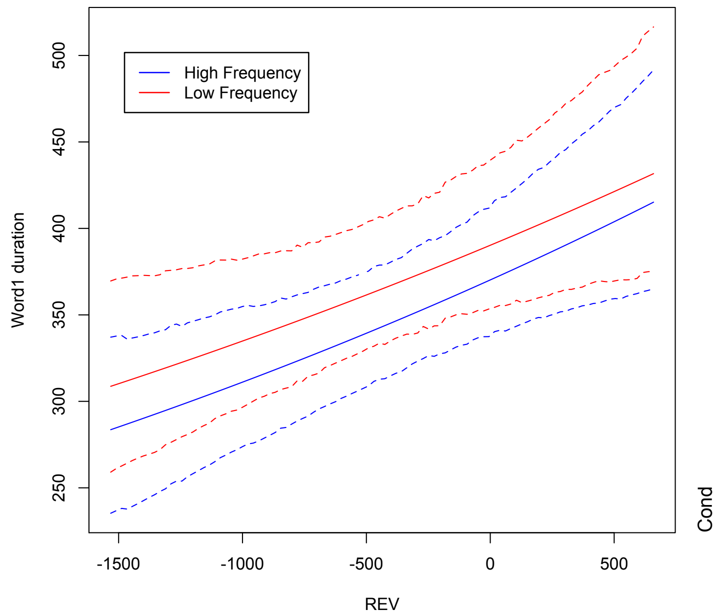


Figure 9. Word2 frequency and REVS predict Word1 duration in Experiment 2. The dotted blue and red lines depict a 95% confidence interval according to MCMC sampling.

Speech onset time analysis

REVS predicted speech onset time. Trials with earlier REVS tended to be trials with overall longer latencies. The final model for log speech onset time duration included random intercepts for items and subjects with uncorrelated random slopes for condition. The model included controls for trial number, error rate by subject, and codability for Word1. The predictors were the REVS and Word2 frequency category. Word2 frequency category did not have a significant effect on the log of speech onset time ($\beta = 0.01344$, $t = 1.28$, $p = 0.19$). In other words, difficulty of Word2 planning did not affect speech onset time. REVS unsurprisingly had a significant effect on the log of speech onset time ($\beta = -0.0002$, $t = -9.7$, $p = < 0.0001$). This suggests that the early looks to the right object were associated with long speech onset time. Speech onset time was 1128.34 longer when speakers looked at the right object the earliest possible as opposed to the latest possible. Fixed effects and their associated significance values can be inspected in table 7. Model-implied lines for the effect of REVS and Word2 frequency on speech onset can be inspected in figure 10.

Table 7

Experiment 2 Speech onset time fixed effect model estimates and significance values after MCMC sampling

	b	*SE*	*t*	*p*
Intercept	2.97200	0.01600	185.78000	0.00000
Trial#	0.00111	0.00036	3.06000	0.00240
Sub error	0.27640	0.10270	2.69000	0.00750
W1H	0.07975	0.03360	2.37000	0.01820
W2freq	0.01344	0.01046	1.28000	0.19990
REVS	-0.00022	0.00002	-9.70000	0.00000
W2freq*REVS	-0.00001	0.00004	-0.33000	0.74190

Note. Trial# refers to the order of trial presentation. Sub error refers to naming error rate by subjects. W1H refers Word1 codability.

Discussion

Experiment 2 replicated Experiment 1 results: low Word2 frequency led to longer Word1 duration than high Word2 frequency did, whereas Word2 frequency did not affect speech onset time. These results further support the role of planning on word duration. Experiment 2 results also provide support for the hypothesis that the timing of Word2 planning affects Word1 duration. Early timing of Word2 planning led to Word1 reduction.

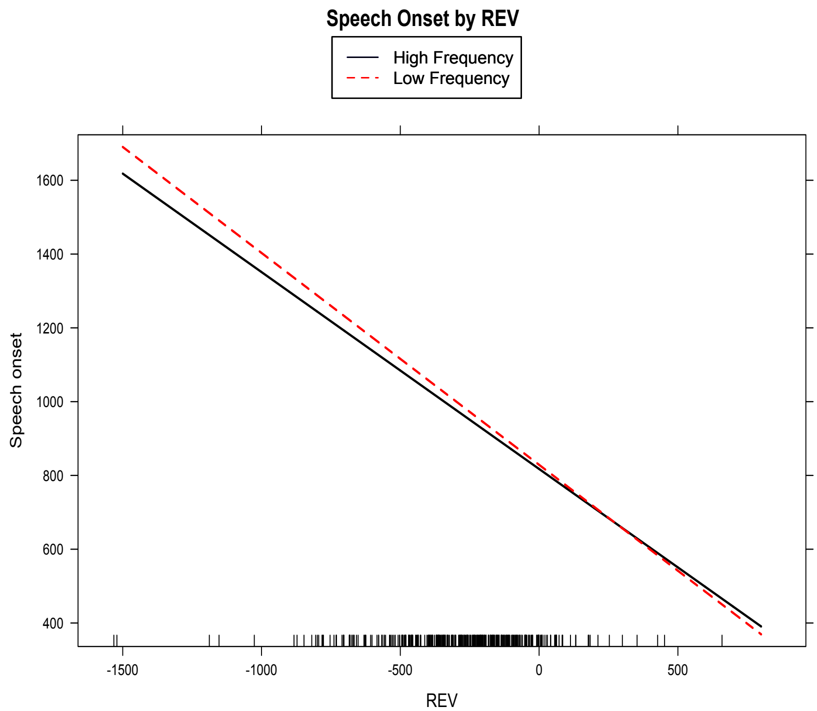


Figure 10. REVS predicts Speech Onset Time in Experiment 2. The black solid line represents the model-implied effect of low frequency and the red dashed line represents the model-implied effect of high frequency. The rug plot on the horizontal axis depicts the distribution of REVS.

The effect of Word2 frequency on Word1 duration suggests that when the total amount of utterance planning increased, Word1 was long. Assuming that the amount of Word1 planning did not change between Word2 frequency conditions, Word1 duration variation should be attributed to planning Word2.

The effect of REVS on Word1 duration showed that in trials with earlier planning, Word1 duration was reduced. Therefore, reduction occurred during trials where Word1 and Word2 planning were closely coordinated. In other words, in certain trials the speaker was able to shift more of Word2 planning prior to speech onset. This resulted in a higher degree of coordination between Word1 and Word2 planning because the speaker had advanced information about planning Word2. As a result the speaker did not require additional time to plan Word2 while articulating Word1. Therefore, instead of slowing down Word1 articulation to allow additional Word2 planning time, the speaker could have sped up Word1 articulation to minimize time on task.

There are also two scenarios where the amount of Word1 planning could have changed on a trial-by-trial basis. One possibility is that reduction occurred in trials where the overall amount of planning both words was reduced. According to this scenario the facility of planning Word1 in addition to Word2 could have contributed to Word1 reduction. This could occur if certain speakers found certain combination of objects easier to name in sequence compared to other combinations of objects. This could have resulted in planning facilitation of Word1, which could have led to its reduction. This account requires the coordination of planning Word1 and Word2. However, this alternate explanation is unlikely because the paired objects were conceptually unrelated to each other.

A different possibility is that reduction occurred in trials where the facility of planning Word1 varied independent of Word2 planning. Even though all left objects had low frequency names, there was still variation in their actual frequency counts and other picture and lexical characteristics. Word1 frequency and all picture characteristics do not appear as a control variables in the word duration model because they were not explaining significant variance in the model. However, Word1 frequency was correlated with Word1 syllable length, Word1 onset phoneme category and item disfluency rate. When excluding those controls, Word1 frequency marginally predicted Word1 duration (t = - 1.79) in addition to Word2 frequency and REVS.

This follow-up analysis should render an account of only Word1 planning influencing Word1 duration less plausible. This is because Word2 frequency and REVS predicted Word1 duration in addition to the marginal effect of Word1 frequency on Word1 duration. If Word1 frequency were solely responsible for Word1 duration variation, Word2 frequency and REVS should have not explained a significant amount of variance in Word1 duration after including Word1 frequency in the model.

As in Experiment1 I conducted a secondary analysis on Word1 duration where I included items with pauses longer than 200 ms. The results showed a significant effect of Word2 frequency and REVS independent of the presence of a pause longer than 200 ms. The table of fixed effects can be inspected in Appendix C. This finding suggests that the effect of planning on Word1 duration could occur independent of disfluent pausing.

In sum, the current experimental results provided further support for the idea that Word2 planning difficulty affected Word1 duration variation. They also provided evidence that the timing of planning affected Word1 duration. Early looks to the right object led to Word1 reduction. In the next two experiments I will directly manipulate factors related to the time available to preplan or plan incrementally.

CHAPTER 4

EXPERIMENT 3: THE EFFECT OF AVAILABLE INCREMENTAL PLANNING TIME ON WORD1 DURATION

In Experiment 3 I further tested the second hypothesis that Word1 duration varies according to the timing of Word2 planning. I manipulated the amount of available incremental planning by varying the presence of the word *and* between the two bare nouns. In two separate blocks speakers either inserted the word *and* (*and* condition) while trying to be fluent, or they tried to not pause when producing two bare nouns (*no pause* condition).

This manipulation is informed by Griffin (2003). She found that Word1 length influenced the timing of Word2 planning only when speakers did not insert the phrase *next to* between naming the two bare nouns of the dual-naming task. This most probably occurred because speakers did not have to estimate the duration of Word1 when two words intervened between the two target words. The additional phrase allowed speakers to always plan incrementally.

The manipulation of the available time for incremental planning is related to the degree of planning coordination. When there is ample time for incremental planning the planning coordination between Word1 and Word2 should be less important. This is because Word2 planning could safely occur completely after speech onset and during articulation of Word1.

If speakers choose to do less preplanning because they know they will have more time available for incremental planning, the *and*/*no pause* manipulation should significantly shift the timing of planning. Word1 duration variation should then only occur in the *no pause* block as opposed to the *and* block. Otherwise, if the timing of Word2 planning is not shifted significantly between the *and* and *no pause* block, word duration variation should still occur according to Word2 frequency independent of the timing manipulation.

Speech onset time should be shorter in the *and* block compared to the *no pause* block. This will support the idea that speakers are shifting their timing of Word2 planning after speech onset. I do not expect an interaction between Word2 frequency and experimental block on speech onset time because Word2 frequency does not affect speech onset time in the current task.

The critical prediction is that Word2 frequency should influence Word1 duration only in the *no pause* condition. This is because the presence of the word *and* should allow for incremental planning without the need to prolong Word1. Whether a main effect of experimental block is also observed is an open question.

Method

Participants

A total of 33 students from the community at UNC Chapel Hill participated in the experiment in exchange for course credit. Nine participants were excluded from the analysis and replaced because of equipment failure ($n = 8$), for not being a speaker of American English ($n = 1$). All other participants were native speakers of American

English or bilinguals of English and another language. The equipment failure rate was due to the experimental software. It overwrote the eyetracking data files for participants with two-digit participant numbers. We discovered the error after running 20 participants. We resumed conducting the experiment within a day of fixing the issue.

Materials and Procedure

Experiment 3 was identical to Experiment 2 except for the addition of a new independent variable. I manipulated the degree of incremental planning independent of Word2 frequency by varying the number of words in the target utterance. In one condition participants were instructed to use two bare nouns when naming the two pictures. In a second condition they were instructed to introduce the connector *and* between the two bare nouns. In both conditions the participants were asked to not pause between words. Participants were exposed to both conditions and in separate blocks. Each block consisted of a single experimental list. Therefore, participants saw both experimental lists and both instruction conditions. The presentation order of the blocks was counterbalanced across participants and across experimental lists. The presentation order of the experimental lists was counterbalanced across participants.

All other equipment, material and procedures were identical to Experiment 2 except for the following changes. I replaced two stimulus pictures that were associated with a high error rate in experiment 2. A list of all the changes in stimuli across all experiments is provided in appendix A. I used the experimental software OpenSesame (Mathot, Schreij, Theeuwes, 2012) to present the experimental stimuli. Audio recording was conducted directly onto the presentation PC via a modified python script incorporated into the experimental software. I used the same condenser microphone as in previous experiments connected to the PC through the Icicle preamplifier made by Blue.

The procedure was different compared to Experiment 2 in the following respects. Before each trial the participants fixated on a dot in the middle of the left side of the screen. I moved the location of the fixation point to match the position used by Griffin (2004) and Meyer et al. (2007). I also added an additional event between the appearance of the fixation point and the onset of the stimuli. The fixation point changed shape when the eyetracker had completed drift correction. The new shape stayed on the screen for 800 ms. I implemented this change to mitigate any issues of difference in level of inter and intra subject preparedness at trial onset. For example, in cases where the duration of the drift correction was prolonged, the participant might not have been adequately prepared for the trial to start. The participant could now anticipate the onset of the trial at the cue of the change in the fixation shape. The participants then named the pictures from left to right and hit the spacebar at the end of the utterance. They then hit the spacebar again to initiate the next trial.

Results

The experimental results show that the *and*/*no pause* manipulation did not affect Word1 duration. However, Word1 duration was longer when Word2 was low frequency rather than high frequency. Also, Word2 frequency did not affect speech onset time. Both of these results are consistent with the previous experimental results. The overall pattern of results based on sample means is depicted in Figure 11. Furthermore, there was an interaction between Word2 frequency and REVS. Word1 duration was reduced when the speaker looked at the right object prior to speech and when Word2 was high frequency (see figure 12). All significant control variables for both Word1 duration and speech onset time are listed in table 8.

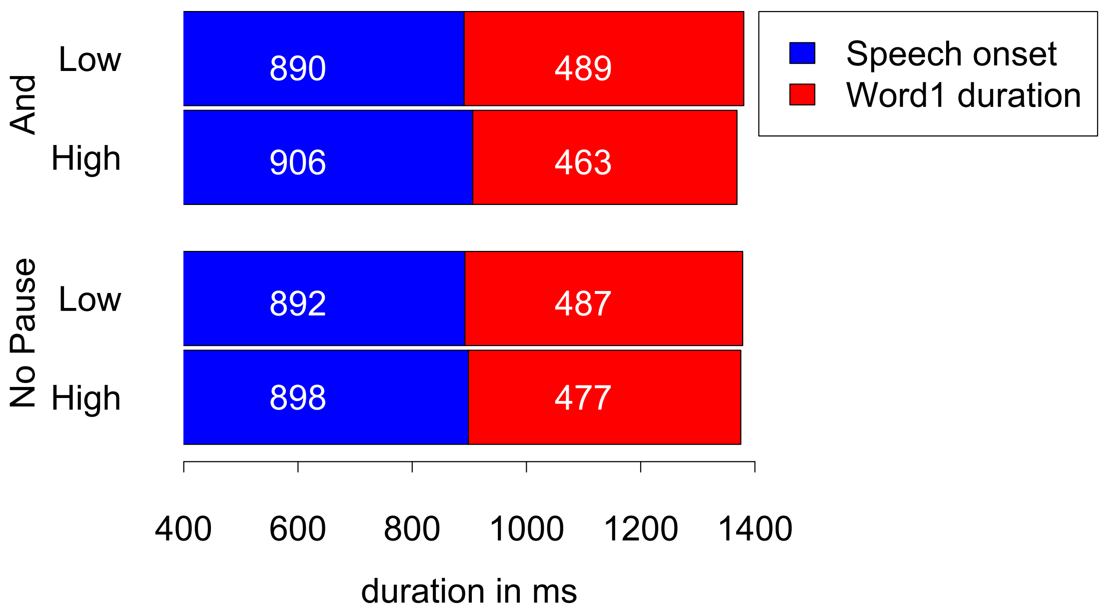


Figure 11. Speech onset time and Word1 duration sample based means according to condition in Experiment 3.

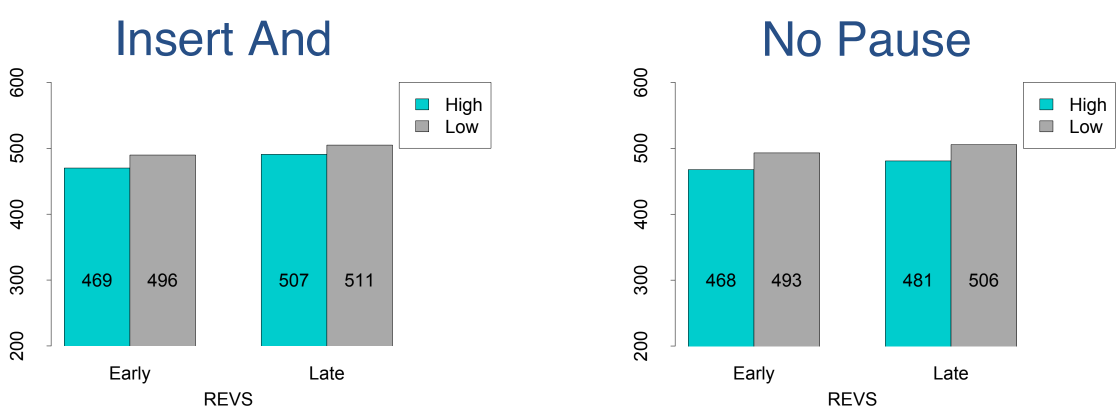


Figure 12. Experiment 3 Word1 duration according the Word2 frequency and REVS. The left panel represents the condition where the word *and* was introduced. The right panel represents the condition where the speaker tried to not pause.

Table 8

Control predictors for each model in Experiment 3

Variable types	Predictors	Word1 duration	Speech onset
Design	Number of syllables	X	
	List		
	Combination of list and Block		X
	Block order	X	X
	Trial	X	
	Pause	X	
Phoneme	Word1 onset	X	X
	Word1 offset	X	
	Word1 offset*Word2 onset		
Performance	Sex		
	Error rate by subject		
	Error rate by item		
	Disfluency rate by subject		
	Disfluency rate by item		
Characteristics	Frequency		
	Word 1 Familiarity		X
	Imageability		
	Word1 and Word 2 Codability		X
	Word1 Visual complexity		

Note. Design variables refer to variables relevant to the design of the experiment. Phoneme variables refer to variables relevant to the onset and offset of Word1 and Word2. Performance variables refer to error and disfluency rate. Characteristics refer Word1 or Word2 picture characteristics.

Exclusions and error rate analysis

The overall error rate before exclusions was 10.24%, or 118 trials out of 1152 trials). Items with a combined naming and accuracy error rate higher than 33% were excluded from further analyses. This led to the exclusion of one item (24 trials or 4.2% of the original data). After the exclusions the error rate was reduced to 8.33% (or 72 trials). I also excluded trials with disfluencies. This led to the exclusion of 18 trials (1.56% of the original data).

Furthermore, I excluded trials with speech onset time longer than 2000 ms (15 trials or 1.3% of the data) and pause duration longer than 200 ms (120 trials or 10.42% of the data). Lastly I excluded four additional trials because of miscellaneous errors. Overall I excluded 24.05% of the data (or 277 trials out of 1104 trials) after excluding one item or 28.22% (or 325 trials out of 1152) before excluding the one item.

I compared the error rate across frequency conditions and type of instructions. The error rate was marginally higher in the low (M = 2.65%, SD = 2.38%) compared to the high (M = 1.52%, SD = 1.76%) frequency condition ($F1$ (1,90) = 2.98, p = 0.08, $F2$ (1, 86) = 0.67, p = 0.41). The error rate did not differ when speakers inserted the word *and* (M = 1.74%, SD = 2.03%) from when speakers were asked not to pause (M = 2.43%, SD = 2.25%) ($F1$ (1,90) = 2.62, p = 0.11, $F2$ (1,86) = 1.14, p = 0.29). There was also no interaction between the frequency condition and the type of instructions ($F1$ (1,90) = 0.04, p = 0.84, $F2$ (1,86) = 0.58, p = 0.81).

I also compared disfluency rate across frequency conditions and type of instructions. The disfluency rate did not differ in the low (M = 0.02%, SD = 0.06%) compared to the high (M = 0.05%, SD = 0.11%) frequency condition ($F1$ (1,90) = 0, p = 0.99, $F2$ (1,86) = 0.54, p = 0.46). The disfluency rate was marginally higher when speakers were asked not to pause (M = 0.05%, SD = 0,11%) from when speakers were asked to insert the word *and* (M = 0.22%, SD = 0.06%) but only by subjects ($F1$ (1,90) = 3.56, p = 0.62, $F2$ (1,86) = 1.54, p = 0.46). There was also a marginal interaction between the frequency condition and the type of instructions by subjects but not by items (*F1* $(1,90) = 3.58$, $p = 0.62$, *F2* $(1,86) = 2.16$, $p = 0.14$).

Eyetracking analysis

The fixation data analysis was the same as in Experiment 2. Speakers showed a preference of looking at the right object after speech onset. On average they looked at the right object after the onset of speech 84% of the time. The average proportion of looking to the right object after speech onset was similar between experimental blocks. In the *and* condition speakers looked at the right object 87% of the time after speech onset. In the *no pause* condition they looked at the right object 80% of the time after speech onset. This suggests that the *and* manipulation induced a similar degree of incremental planning compared to the *no pause* condition.

Word1 duration analysis

REVS and Word2 frequency influenced word duration. REVS also interacted with Word2 frequency. However, the *and*/*no pause* manipulation did not influence Word1 duration. The final model for log word duration included random intercepts for items and subjects with uncorrelated random slopes for condition and experimental block. The model included controls for number of syllables, trial number, block order, pause duration, phoneme category of Word1 onset and offset. The predictors were the *and*/*no pause* manipulation (Experimental block), REVS and Word2 frequency category. Word2 frequency category had a significant effect on the log of Word1 ($\beta = 0.0163$, $t = 2.26$, $p < 0.05$). To interpret model-implied effects I back-transformed the data to the original scale. Word1 duration was 16 ms longer when Word2 was low frequency as opposed to high frequency. REVS[2] had a significant effect on the log of Word1 duration (β = 0.0001, t = 3.85, p < 0.001). Word1 duration was 133 ms shorter when REVS was at its maximum negative value as opposed to its maximum positive value. In other words, the earlier looks to the right object with relation to speech onset the more reduced Word1 duration. However, there was a significant interaction between REVS and Word2 frequency category (β = -0.0001, t = -2.03, p = 0.05). The interaction

Table 9

Experiment 3 Word1 duration fixed effect model estimates and significance values after MCMC sampling

	b	*SE*	*t*	*p*
Intercept	2.67900	0.01603	167.12000	0.0000
Pause	0.00629	0.00316	1.99000	0.0467
Syllables	0.03057	0.01737	1.76000	0.0787
First	-0.02965	0.01004	-2.95000	0.0032
Trial#	0.00079	0.00024	3.30000	0.0010
W1onset	0.02567	0.00712	3.61000	0.0003
W1offset	0.01310	0.00484	2.71000	0.0070
W2freq	0.01634	0.00724	2.26000	0.0242
REV	0.00006	0.00001	3.85000	0.0001
Inst	0.00119	0.01113	0.11000	0.9153
W2freq*REV	-0.00004	0.00002	-2.03000	0.0428

Note. First refers to experimental block order. Trial# refers to the order of trial presentation. W1onset refers to initial Word1 phoneme category. W1offset refers to final Word1 phoneme category. Inst refers to the experimental block of inserting and or not pausing.

[2] The effect of REVS also predicts duration when REVS is treated as a categorical variable of looks before vs. after the onset of speech.

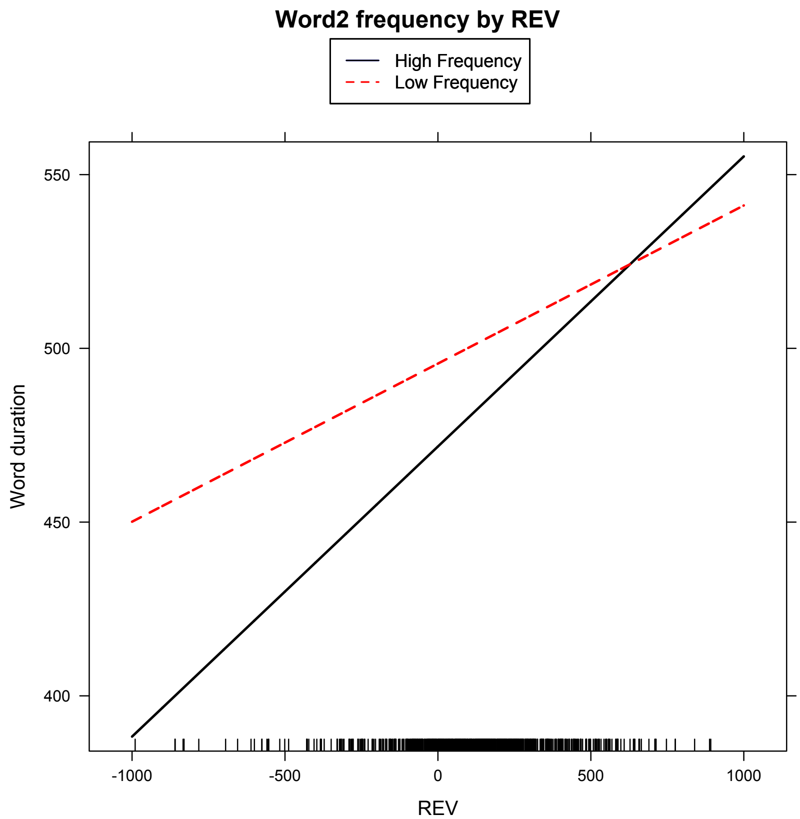


Figure 13. The interaction between Word2 frequency and REVS predict Word1 duration in Experiment 3. The black solid line represents the model-implied effect of low frequency and the red dashed line represents the model-implied effect of high frequency. The rug plot on the horizontal axis depicts the distribution of REVS

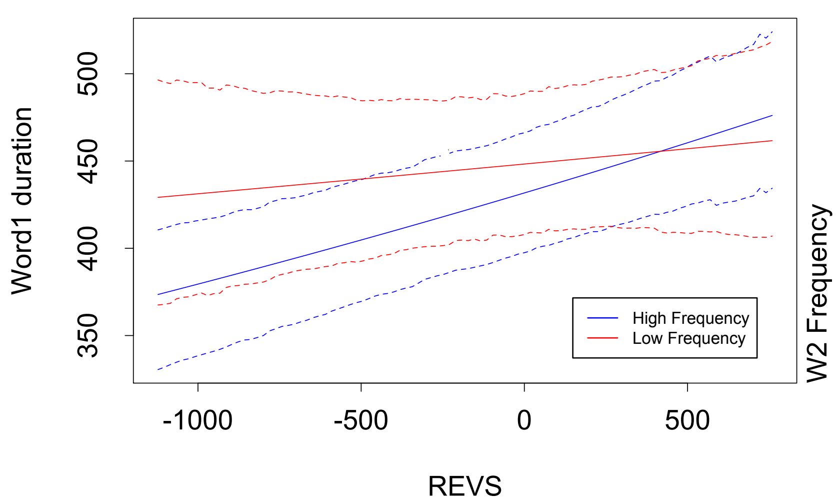


Figure 14. The interaction between Word2 frequency and REVS predict Word1 duration in Experiment 3. The blue line represents high Word2 frequency and the red line low Word2 frequency. The dotted blue and red lines depict a 95% confidence interval according to MCMC sampling.

suggests that Word1 duration was long independent of Word2 frequency category when the speaker looked at the right object after the onset of speech. However, Word1 duration was reduced when Word2 frequency was high and the speaker looked at the right object before speech onset. Fixed effects and their associated significance values can be inspected in table 9. The pattern of results can be inspected in figures 13 and 14.

Speech onset time analysis

REVS and the interaction between REVS and Word2 frequency influenced speech onset time. However, the *and/no pause* manipulation did not influence speech onset time. The final model for log speech onset time duration included random intercepts for items and subjects. All slopes for Word2 frequency and experimental block were uncorrelated. The model included controls for block order, the combination of block and experimental list, Word1 onset phoneme category, codability for Word1 and Word2, Word1 familiarity. The predictors were the REVS and Word2 frequency category. Word2 frequency category did not have a significant effect on the log of speech onset time (β = 0.0033, t= 0.64, p = 0.52). In other words, difficulty of Word2 planning did not affect speech onset time. REVS unsurprisingly had a significant effect on the log of speech onset time (β = -0.0001, t = -7.03, p = < 0.001). This suggests that when speakers looked at the right object before speech onset, speech onset time was longer. Speech onset time was 242 ms longer when speakers looked at the right object before speech onset as opposed to after speech onset. REVS also interacted with Word 2 frequency (β = -0.0001, t = -2.04, p = < 0.05). Fixed effects and their associated significance values can be inspected in table 10. The pattern of results can be inspected in figures 15 and 16.

Table 10

Experiment 3 Speech onset time fixed effect model estimates and significance values after MCMC sampling

	b	*SE*	*t*	*p*
Intercept	2.95400	0.01336	221.13000	0.00000
Version	-0.02031	0.00998	-2.04000	0.04190
First	-0.05390	0.00805	-6.69000	0.00000
W1fam	-0.00928	0.00765	-1.21000	0.22570
W1H	0.07738	0.02692	2.87000	0.00420
W2H	0.01856	0.00835	2.22000	0.02600
W2freq	0.00335	0.00522	0.64000	0.52010
REV	-0.00010	0.00001	-7.03000	0.00000
Inst	0.01154	0.00833	1.39000	0.16620
W2freq*REVS	-0.00004	0.00002	-2.04000	0.04180
Inst*REVS	0.00004	0.00003	1.55000	0.12040

Note. Version refers to the combination of experimental list and experimental block. W1fam refers to Word1 familiarity. W1H and W2H refer to Word1 and Word2 codability. Inst refers to the experimental block of inserting and or not pausing.

Discussion

Experiment 3 successfully replicated the effect of Word2 frequency and REVS on Word1 duration. Word1 duration was longer when followed by a low frequency Word2 as opposed to a high frequency Word2. Also, early REVS led to a reduction of Word1 duration. Therefore I again found support for the hypotheses that both planning difficulty and the timing of planning affects word duration.

Furthermore, Experiment 3 provided evidence of an interaction between the timing of planning and Word2 frequency. The interaction occurred because of trial-by-trial variation in the timing of planning with relation to the onset of speech as opposed to the independent manipulation of Word2 timing of planning.

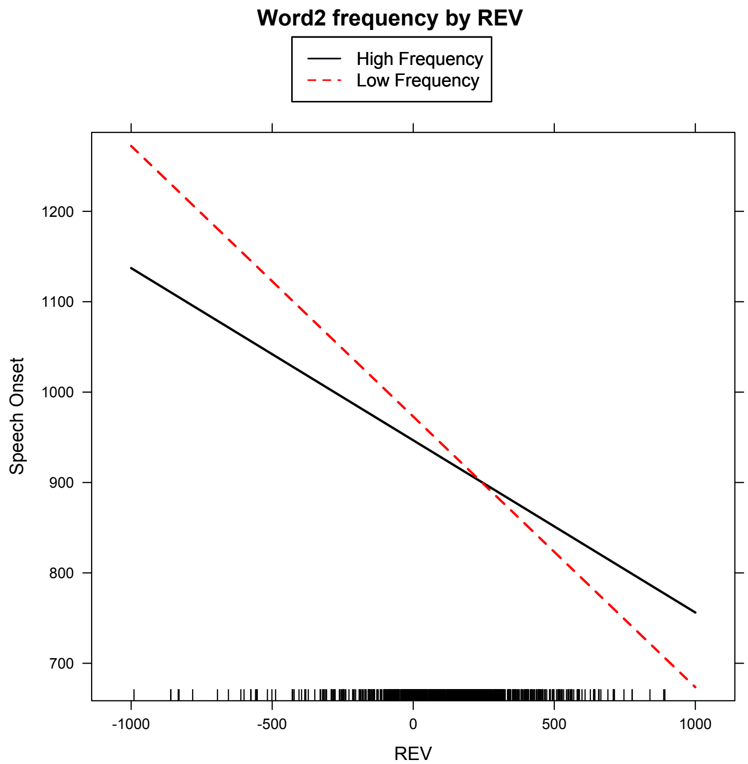


Figure 15. The interaction between Word2 frequency and REVS predict speech onset time in Experiment 3. The black solid line represents the model-implied effect of low frequency and the red dashed line represents the model-implied effect of high frequency. The rug plot on the horizontal axis depicts the distribution of REVS.

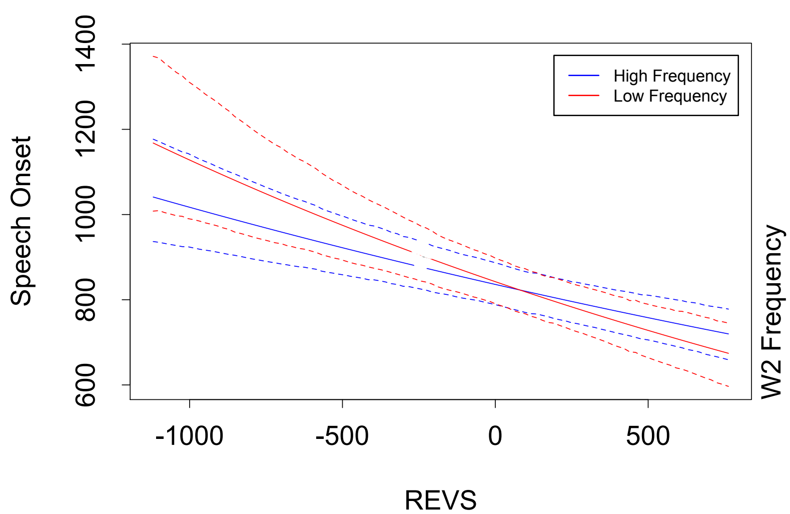


Figure 16. The interaction between Word2 frequency and REVS predict speech onset time in Experiment 3. The blue line represents high Word2 frequency and the red line low Word2 frequency. The dotted blue and red lines depict a 95% confidence interval according to MCMC sampling.

The *and*/*no pause* manipulation led to a change in the proportion of trials where speakers fixated the right object prior to speech onset. As predicted, speakers looked less frequently at the right object prior to speech onset when they inserted the word *and* between the two target words. However, this change in the REVS did not significantly affect Word1 duration. This potentially occurred because the rate of incremental planning was overall high in the task. The interpretation that the degree of incremental planning did not change is supported by the lack of a significant effect of the insertion of *and* on speech onset time.

Despite the lack of a difference according to the *and*/*no pause* manipulation in the degree of incremental planning, there is some evidence that planning coordination was an important part of word duration variation. When REVS occurred early with relation to speech, Word1 duration was reduced when Word2 was high frequency. However, when REVS occurred late with relation to speech Word1 duration was long independent of Word2 frequency. This finding suggests that high frequency words can ease coordination of planning because they require less amount of planning. Supporting evidence comes from the fact that speech onset time only varied according to the interaction between Word2 frequency and REVS. Given that speech onset time did not differ according to Word2 frequency this result suggests that a high frequency Word2 required less preplanning time compared to a low frequency Word2.

The main effect of REVS and its interaction with Word2 frequency were no longer significant when I treated REVS as categorical (before vs. after speech onset). This is potentially the case because speakers had a general tendency to look at the right object after speech onset. In a secondary analysis, to demonstrate that looks to the right object prior to speech onset were influential on the interaction I excluded all looks prior to speech onset. The interaction term was now marginal.

Experiment 2 results suggested that Word1 duration variation might have not been solely due to Word2 planning difficulty. I therefore outlined three possible scenarios of how Word1 duration could have varied. The first and second scenario required the coordination of Word1 and Word2 planning. However, the second scenario is highly unlikely given the experimental design. I will therefore not discuss it any further. The third scenario required only variation in Word1 planning. According to the presence of an interaction between Word2 frequency and REVS I can safely exclude the third scenario.

I also conducted a secondary analysis where I included pauses longer than 200ms for the word duration model. As in Experiment 2 the results showed that both effects of REVS and Word2 frequency were significant. You can find the fixed effects table for the model in appendix D.

CHAPTER 5

EXPERIMENT 4: THE EFFECT OF AVAILABLE PREPLANNING TIME ON WORD1 DURATION

In Experiment 4 I varied the available preplanning time. I used a manipulation that required participants to start speaking before a specific deadline. In two separate blocks the speakers had to start speaking either before a 2000 ms or a 960 ms deadline. 2000 ms was the maximum allowable speech onset time for inclusion in data analysis in all previous experiments. 960 ms was the average speech onset time in Experiment 1.

The speakers received feedback on their performance at the end of each trial. This way the speaker had to figure out a specific strategy to meet the demands of the task. If they failed to meet task requirements, they received negative feedback at each trial. Thus participants could keep adjusting their performance on a trial-by-trial basis.

The deadline manipulation is informed by previous research in incremental planning (Ferreira & Swets, 2002). Also Damian (2003) and Kello et al. (2000) used the deadline manipulation to test their hypotheses about the relationship between planning and word duration.

Predictions for the current experiment are very similar to Experiment 3. Here I will only focus on predictions that might differ between the two experiments. Based on the incremental planning literature I would expect speech onset time to be short in the early deadline condition and long in the late deadline condition. However, according to Kello et al. (2000) if the speaker encounters heightened cognitive load the early deadline condition should lead to longer speech onset compared to the late deadline condition. On the other hand Damian (2003) but also Damian and Dumay (2007) suggest that time pressure should have no effect on the degree of incremental planning.

According to the incremental planning literature, Word1 duration could potentially be longer in the early compared to the late deadline condition. However, there is an alternate possibility that Word1 duration will be reduced in the early compared to the late deadline condition. This is predicted by Ferreira & Swets (2002) who found that utterance duration was overall reduced when speakers had to meet a deadline to begin speaking. This outcome suggests that speakers are overall planning faster when they need to meet a deadline. As in all previous experiments, I expect low frequency rather than high frequency Word2 to lead to longer Word1 duration.

Method

Participants

A total of 34 students from the community at UNC Chapel Hill participated in the experiment in exchange for course credit. Ten participants were excluded from analyses and replaced because of equipment failure (eyetracking failure $n = 3$, voice key malfunction, $n = 6$), because the lab assistant forgot to turn on the microphone ($n = 1$). All other participants were native speakers of American English or bilinguals of English and another language. The equipment failure rate was due to either failure with the eyetracking system ($n = 3$) or because of voice key malfunction ($n = 6$).

Materials and Procedure

Experiment 4 was identical to experiment 3 except for a change in the Word2 timing manipulation. Participants were instructed to speak before a deadline. In the early deadline condition speakers were supposed to start speaking within 960 ms after the onset of the stimuli. In the late deadline condition speakers were supposed to start speaking within 2000 ms after the onset of the stimuli. In both conditions the participants were asked to not pause between words. Participants were exposed to both conditions and in separate blocks. Each block consisted of a single experimental list. Therefore, participants saw both experimental lists and both instruction conditions. The presentation order of the blocks was counterbalanced across participants and across experimental lists. The presentation order of the experimental lists was counterbalanced across participants.

All equipment, material and procedures were identical to Experiment 3 except for the following adjustments in the procedure. Participants were instructed to try not to pause between the two words but also to try and speak as soon as the stimuli appeared on the screen, without sacrificing accuracy. At the end of each trial participants received feedback on whether they spoke early enough to beat the deadline. They received the message "Good!" if they were successful. Otherwise, they received the message "Please start speaking earlier." At the beginning of each block participants were warned that the deadline would either be easy or difficult to beat. As in all previous experiments participants received 12 practice trials before starting each experimental block.

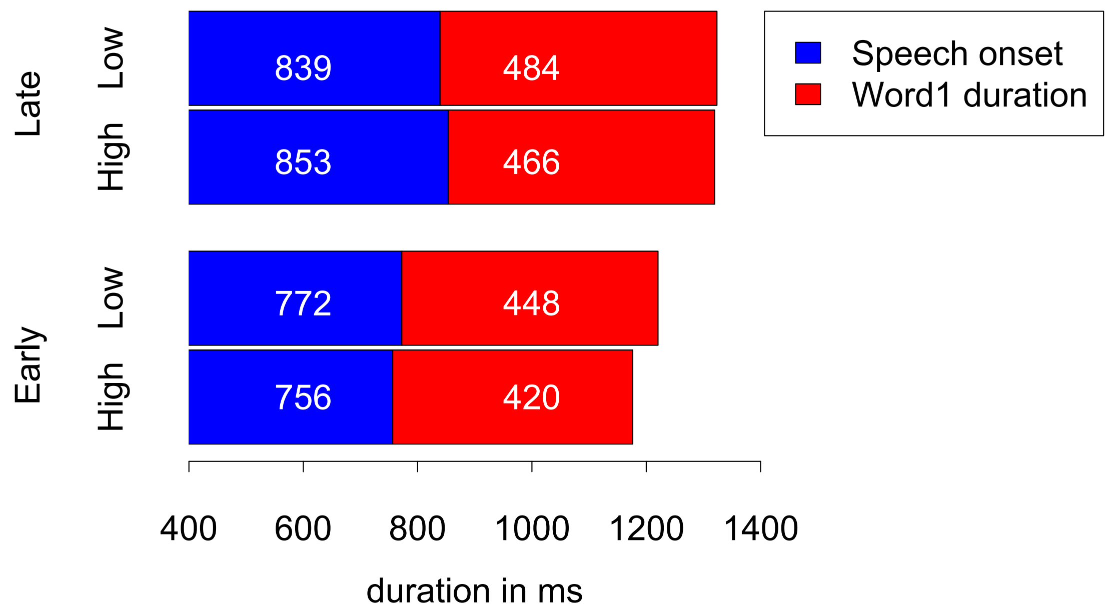


Figure 17. Experiment 4 Speech onset time and Word1 duration sample based means according to condition.

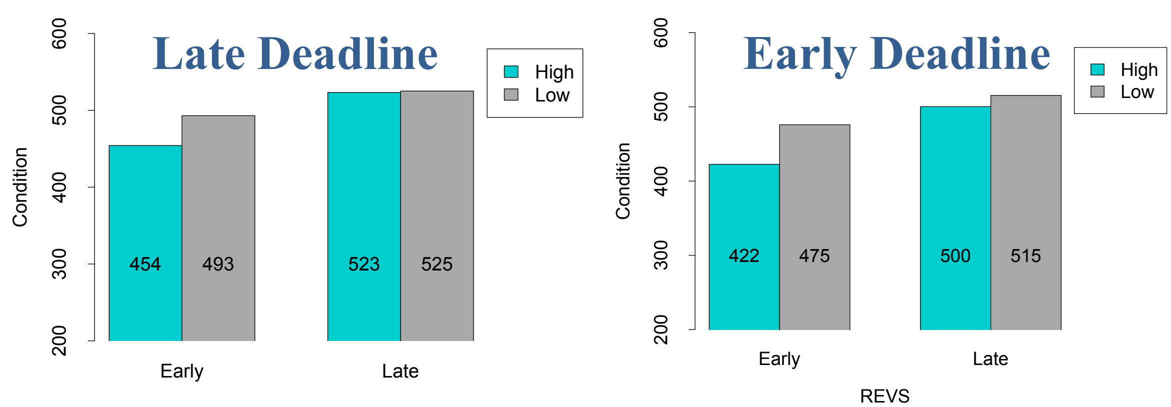


Figure 18. Experiment 4 Word1 duration according the Word2 frequency and REVS. The left panel represents the late deadline condition. The right panel represents the early deadline condition.

Table 11

Control predictors for each model in Experiment 4.

Variable types	Predictors	Word1 duration	Speech onset
Design	Number of syllables	X	
	List		
	Combination of list and Block		
	Block order	X	X
	Trial		
	Pause	X	X
Phoneme	Word1 onset	X	
	Word1 offset	X	
	Word1 offset*Word2 onset		
Performance	Sex		
	Error rate by subject		
	Error rate by item		
	Disfluency rate by subject		
	Disfluency rate by item		
Characteristics	Frequency		
	Word 1 Familiarity		
	Imageability		
	Word1 Codability		X
	Word1 Visual complexity		

Note. Design variables refer to variables relevant to the design of the experiment. Phoneme variables refer to variables relevant to the onset and offset of Word1 and Word2. Performance variables refer to error and disfluency rate. Characteristics refer Word1 or Word2 picture characteristics.

Results

The results showed that Word1 duration was longer when Word2 was low frequency rather than high frequency. However, Word2 frequency did not affect speech onset time. The pattern of results is depicted in Figure 17. Both of these results are consistent with all previous experimental results. Contrary to Experiment 3 there is evidence that the timing manipulation affected word duration. Word1 duration was shorter in the early deadline compared to the late deadline condition. Furthermore, there was an interaction between Word2 frequency and REVS. Word1 duration was reduced in duration when the speaker looked at the right object prior to speech onset and when Word2 frequency was low (see figure 18). All significant control variables for both Word1 duration and speech onset time are listed in table 11.

Error rate analysis

The overall error rate before exclusions was (16.06% or 185 trials out of 1152 trials). Items and participants with a combined naming and accuracy error rate higher than 33% were excluded from further analyses. This led to the exclusion of two items (96 trials or 8.3% of the original data) and the replacement of one participant. After the exclusions the error rate was reduced to 13.07% (or 138 trials out of 1056).

The error rate was higher in the low frequency condition ($M = 0.17$, $SD = 0.12$) compared to the high frequency condition ($M = 0.11$, $SD = 0.09$) in the by subjects analysis $F1$ $(1,91) = 7.16$, $p < 0.01$ but only marginal in the by items analysis $F2$ $(1,83) = 3.12$, $p = 0.08$.

The disfluency rate was 3.5% before exclusions (40 trials). After exclusions it decreased to 3.12% (or 33 trials). There was no difference in disfluency rate between deadline condition or depending on frequency conditions. However, there was a significant interaction by subjects and marginally significant by items between the two factors ($F1$ $(1,91) = 3.75$, $p = 0.055$, $F2$ $(1,83) = 4.41$, $p < 0.05$). The difference in disfluency rate between the two deadline conditions was larger in the high frequency compared to the low frequency condition.

Speakers also encountered a more difficult time meeting the deadline in the early deadline compared to the late deadline condition. You can inspect the success rate in meeting the deadline in the following table.

Table 12**.** Success rate in speaking prior to deadline in Experiment 4

	Early	Late
High	69%	99%
Low	66%	100%

I further excluded: (1) latencies longer than 2000 ms (0.06% or 7 trials), (2) trials where speakers did not look at the right object before looking at the left object (1.3% or 13 trials), (3) pauses longer than 200 ms (17.14% or 181 trials). The overall rate of exclusion was 37% after excluding the two items (or 395 trials out of 1056) or 43% without excluding the two items (or 491 trials out of 1152)

Eyetracking analysis

The fixation data analysis was the same as in Experiment 2. Speakers showed a preference of looking at the right object before the onset of speech. On average they looked at the right object before the onset of speech 85% of the time. You can consult table 13 for a breakdown according to conditions. I then used this measure to predict Word1 duration and speech onset time.

Table 13

Proportion of looks to the right object prior to speech onset in Experiment 4

	Early	Late
High	88%	88%
Low	88%	83%

Word1 duration analysis

REVS, the deadline condition and Word2 frequency influenced word duration. REVS also interacted with Word2 frequency but not with the deadline condition when influencing word duration. The final model for log word duration included random intercepts for items and subjects with uncorrelated random slopes for condition and deadline condition. The model included controls for number of syllables, block order, pause duration, phoneme category of Word1 onset and offset. The predictors were the REVS and Word2 frequency category. Word2 frequency category had a significant effect on the log of Word1 ($\beta = 0.0361$, $t = 4.18$, $p < 0.0001$). To interpret model-implied effects I back-transformed the data to the original scale. Word1 duration was 37 ms longer when Word2 was low frequency as opposed to high frequency. REVS[3] had a significant effect on the log of Word1 duration ($\beta = 0.0001$, $t = 3.09$, $p < 0.05$). Word1 duration was 85 ms shorter when REVS was at its maximum positive value as opposed to its maximum negative value. In other words, the earlier looks to the right object with relation to speech onset the more reduced Word1 duration. There was also a significant interaction between REVS and Word2 frequency category ($\beta = -0.0001$, $t = -2.06$, p < 0.05). The interaction suggests that Word1 duration was long independently of Word2

[3] REVS had a significant effect on Word1 duration even when treated as a categorical variable (looks before vs. looks after speech onset).

frequency category when the speaker looked at the right object after the onset of speech. However, Word1 duration was reduced when Word2 frequency was high and the speaker looked at the right object before speech onset. Fixed effects and their associated significance values can be inspected in table 14. The pattern of results can be inspected in figures 19 and 20.

Table 14

Experiment 4 Word1 duration fixed effect model estimates and significance values after MCMC sampling

	b	*SE*	*t*	*p*
Intercept	2.6580	0.0148	179.2600	0.0000
Pause	0.0074	0.0032	2.3000	0.0216
First	-0.0100	0.0078	-1.2800	0.2013
Syllables	0.0502	0.0175	2.8800	0.0042
W1onset	0.0230	0.0072	3.2000	0.0015
W1offset	0.0165	0.0049	3.3500	*0.0009*
Instr	-0.0298	0.0079	-3.7700	0.0002
W2freq	0.0361	0.0087	4.1800	0.0000
REVS	0.0001	0.0000	3.9000	0.0001
W2freq*Instr	0.0152	0.0096	1.5800	0.1156
W2freq*REVS	-0.0001	0.0000	-2.0600	0.0398
Pause*Instr	0.0118	0.0052	2.2800	0.0229
Pause*Syllables	0.0130	0.0053	2.4700	0.0137
Pause*Instr*REVS	-0.0001	0.0000	-2.9400	0.0034
Pause*W2freq*REVS	-0.0001	0.0000	-1.9000	0.0578
Instr*W2freq*REVS	-0.0327	0.0195	-1.6800	0.0935

Note. Instructions refers to the experimental block of inserting and or not pausing. First refers to the block order. Trial# refers to the order of trial presentation. W1on refers to Word1 onset phoneme category. W1off refers to Word1 offset phoneme category. Feed refers to accuracy for beating the speech onset deadline.

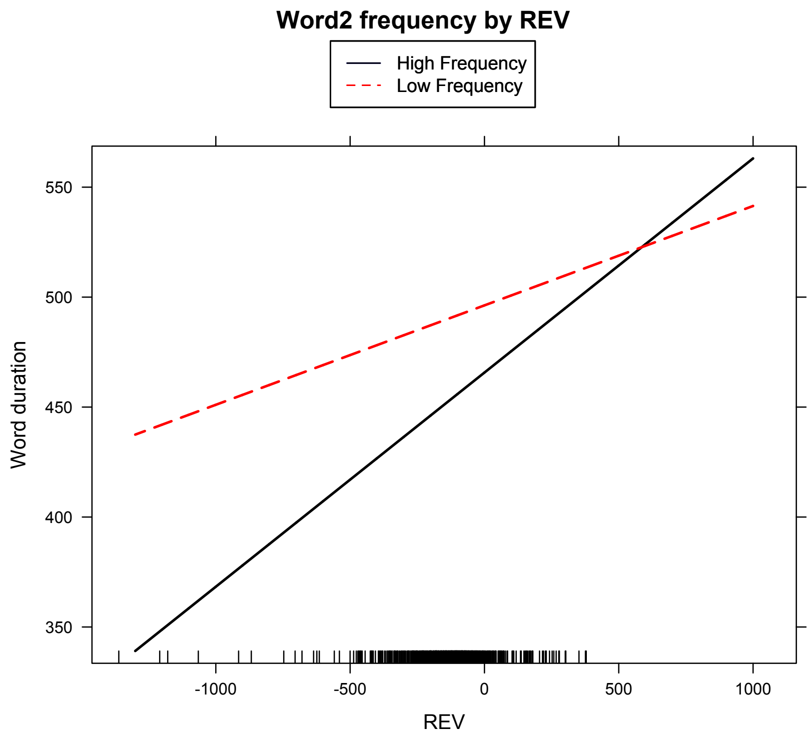


Figure 19. The interaction between Word2 frequency and REVS predict Word1 duration in Experiment 4. The black solid line represents the model-implied effect of low frequency and the red dashed line represents the model-implied effect of high frequency. The rug plot on the horizontal axis depicts the distribution of REVS.

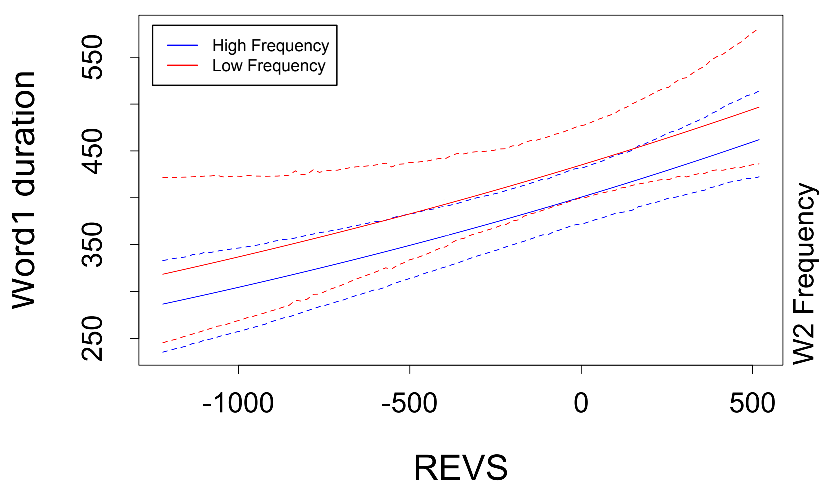


Figure 20. The interaction between Word2 frequency and REVS predict Word1 duration in Experiment 4. The blue line represents high Word2 frequency and the red line low Word2 frequency. The dotted blue and red lines depict a 95% confidence interval according to MCMC sampling.

Speech onset time analysis

The final model for log speech onset time duration included random intercepts for items and subjects with uncorrelated random slopes for deadline and subject intercepts. The model included controls for block order, codability for Word1 and pause duration. The predictors were the deadline condition, REVS and Word2 frequency category. I also included interaction terms between the three predictor variables and the trial-by-trial feedback speakers received regarding their performance meeting the deadline. Word2 frequency category did not have a significant effect on the log of speech onset time ($\beta = 0.004$, $t = 0.6$, $p = 0.55$). In other words, difficulty of Word2 planning did not affect speech onset time. REVS unsurprisingly, had a significant effect on the log of speech onset time ($\beta = -0.0001$, $t = - 6.79$, $p = < 0.0001$). This suggests that when speakers looked at the right object before speech onset, speech onset time was longer. Speech onset time was 861 ms longer when speakers looked at the right object before as opposed to after speech onset. There was also a significant effect of the deadline condition ($\beta = -0.0227$, $t = -2.55$, $p < 0.05$). In the late deadline condition, speakers initiated their utterances 73 ms later compared to the early deadline condition. Fixed effects and their associated significance values can be inspected in table 16. The pattern of results can be inspected in figure 21.

Discussion

Experiment 4 yet again replicated the effect of Word2 frequency and REVS on Word1 duration. Word1 duration was longer when followed by a low frequency Word2 as opposed to a high frequency Word2. Also, early REVS led to a reduction of Word1 duration. Lastly, there was an interaction between REVS and Word2 frequency. Word1

Table 15

Experiment 4 Speech onset time fixed effect model estimates and significance values after MCMC sampling

	b	*SE*	*t*	*p*
Intercept	2.8840	0.0126	229.1400	0.0000
Pause	0.0026	0.0030	0.8700	0.3865
First	-0.0443	0.0070	-6.3200	0.0000
W1H	0.0970	0.0288	3.3700	0.0008
Instr	-0.0227	0.0089	-2.5500	0.0110
W2frequency	0.0040	0.0066	0.6000	0.5491
REVS	-0.0001	0.0000	-6.7900	0.0000
Feed*Instr	0.2808	0.0185	15.1900	0.0000
Feed*W2freq	-0.0019	0.0133	-0.1400	0.8892
Feed*REVS	0.0001	0.0000	3.0600	0.0023
W2Freq*REVS	0.0000	0.0000	0.1100	0.9103
Feed*W1Freq*REVS	-0.0002	0.0001	-2.4100	0.0165

Note. Instr refers to the experimental block of inserting and or not pausing. First refers to the block order. First refers to the block order. W1H refers to Word1 codability. Feed refers to accuracy for beating the speech onset deadline.

duration was reduced when looks to the right object occurred prior to speech onset and when Word2 frequency was high. Therefore I again found support for the hypothesis that the timing of planning affects word duration.

As expected, speakers took less time to begin speaking in the early as opposed to the late condition. However, this didn’t appear to require speakers to do additional incremental planning, which would have resulted in a longer Word1 duration for the early deadline condition. Instead, the early deadline also led to shorter Word1 durations. Speakers probably increased their rate of processing in the early deadline condition as opposed to the late deadline condition. This is consistent with Ferreira & Swets’ (2003) results. Importantly though, the effect of Word2 frequency on Word1 duration still held independent of the effect of the deadline condition.

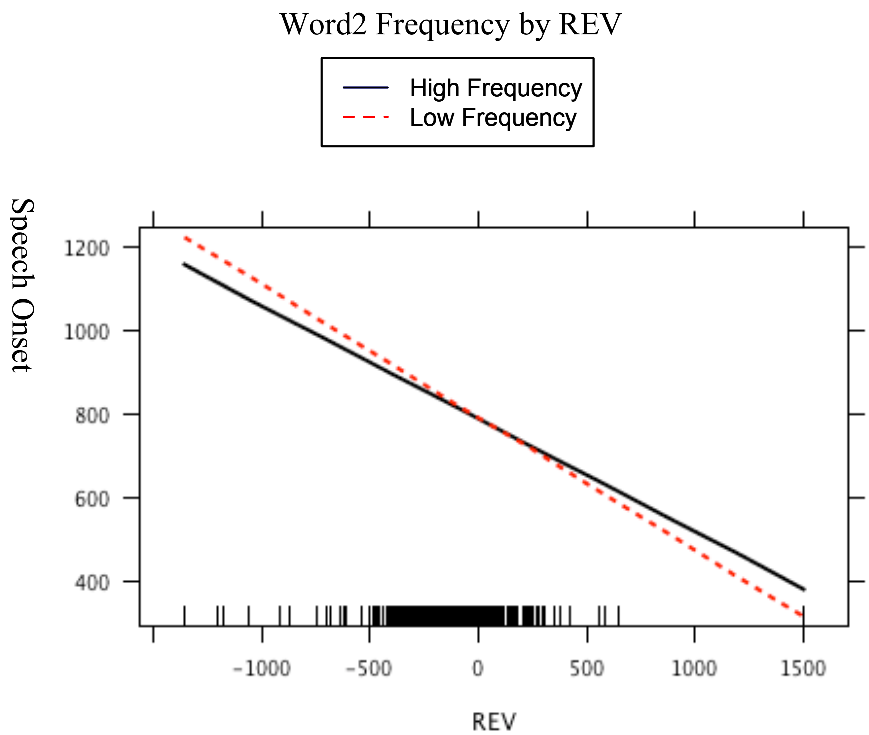


Figure 21. REVS predicts speech onset time in Experiment 4. The black solid line represents the model-implied effect of high frequency and the red dashed line represents the model-implied effect of low frequency. The rug plot on the horizontal axis depicts the distribution of REVS

As in Experiment 3, REVS interacted with Word2 frequency when influencing Word1 duration. When speakers looked at the right object early with relation to speech onset Word1 duration was reduced when Word2 was high frequency. However, when speakers looked at the right object late with relation to speech onset Word1 duration was overall long independent of Word2 frequency.

The effects of REVS on Word1 duration were robust even when treating REVS as a categorical variable (looks to the right object before vs. after speech onset). This finding is consistent with Experiment 3 results and further supports the idea that the onset of speech is a critical point of how the timing of planning influences word duration. Words are reduced when looks to the right object occur prior to speech onset.

However, the interaction between REVS and Word2 frequency did not remain significant when treating REVS as categorical. This finding is again consistent with Experiment 3 results. To show that speech onset was influential for the interaction I did a follow up analysis. I excluded all trials with looks that occurred after speech onset (as in Experiment 3). The interaction was not significant anymore.

Similarly to Experiment 3 there was also an interaction between Word2 frequency and REVS when predicting Speech onset time. However this relationship occurred only when speakers were successful at beating the deadline to begin speaking.

As in all previous experiments, I conducted a secondary analysis where I also included trials with pauses longer than 200 ms. All previously significant predictors were still significant and independent of an interaction with a pause longer than 200 ms. The fixed effect for this model can be consulted in Appendix E.

CHAPTER 6

GENERAL DISCUSSION

In four different experiments speakers produced two-word utterances e.g., *skunk hand*. The frequency of the second word and the timing of planning the second word affected the duration of the first word. Low Word2 frequency led to long Word1 duration whereas high Word2 frequency led to short Word1 duration. Furthermore, early looks with relation to speech onset reduced Word1 duration. Late looks with relation to speech onset led to long Word1 duration.

Even more interestingly, in the last two experiments the timing of Word2 planning interacted with Word2 frequency. Early looks to the right object with relation to speech onset reduced Word1 duration when Word2 frequency was high. However, Word2 frequency did not affect Word1 duration when looks to the right object occurred late with relation to speech onset. In this case Word1 duration was overall long. This suggests that speakers can reduce Word1 duration when they have advance information about Word2 planning prior to speech onset. When the speakers estimate that Word2 is easy to plan prior to speech onset, they can speed up Word1 articulation. The speed up could be driven by a reduced need to plan Word2 while articulating Word1, presumably because of early Word2 timing of planning. Otherwise, it could be driven by the facility of planning Word2 while producing Word1. In other words, Word1 reduction could occur because of preplanning facilitation or incremental planning facilitation.

My results are consistent with other studies that suggest planning affects word duration. There is substantial evidence that different types of predictability influence word duration (e.g., Aylett & Turk, 2004; Bell et al., 2009; Gahl & Garnsey, 2004; Fowler & Housum, 1987; Kahn & Arnold, in press; Lieberman, 1963; Watson, Arnold, & Tanenhaus, 2008). There is also evidence for the role of planning difficulty on word duration (e.g., Bell et al., 2003; Christodoulou, 2009; Clark & Fox Tree, 2002; Fox Tree & Clark, 1997). However, none of these studies can safely conclude that planning has a direct effect on word duration variation.

Word duration can cause variation in the timing of planning (Griffin, 2003 but also see Meyer et al., 2007). The current experimental results suggest that the inverse relationship also holds true. Variation in the timing of planning can cause variation in word duration. This finding could potentially inform the literature on the flexibility of the scope of planning on possible ways that planning variation can affect speech apart from speech onset time (e.g., Schriefers & Teruel, 1999; Wagner et al., 2010).

The current experimental results could also have implications for research on incremental planning. There is some evidence that incremental planning affects utterance duration (Ferreira & Swets, 2002). However utterance duration is confounded with pause duration. If planning upcoming information influences word duration, the question that arises is how does incremental planning affect word duration compared to pausing.

In the current data, word duration variation always covaried with pausing. In all analyses I controlled for the possible effect of pausing on word duration and therefore the effect of planning on word duration is independent of the effect of pausing.

Furthermore, the independent contribution of each effect on word duration held even when including disfluent pauses (pauses longer than 200 ms). These results suggest that incremental planning could potentially have independent effects on word duration and pausing. In other words speakers were not prolonging their words merely to avoid pausing. They were varying word duration according to how much information they had already planned at the moment they started producing a particular word.

The effort to avoid pausing could however have a separate contribution on word duration. This is because there was variation in meeting this requirement according to condition. On average the rate of pausing longer than 200 ms was higher when Word2 was low frequency as opposed to high frequency. This pattern is depicted in Figure 22.

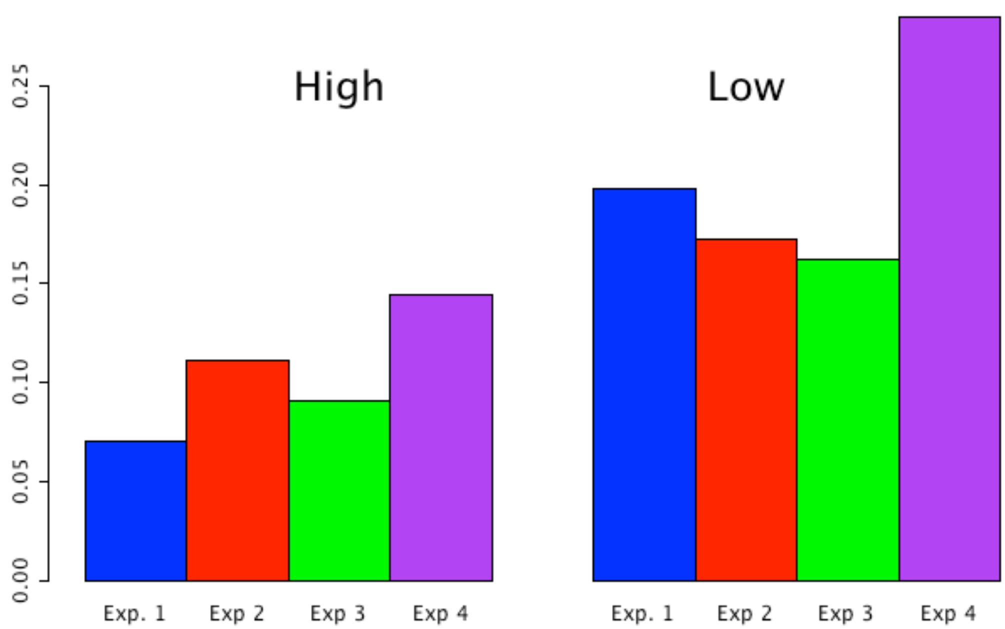


Figure 22. Proportion of trials in which the participant paused longer than 200 ms across all experiments, in each Word2 frequency condition, for each experiment.

This difference in performance did not drive any of the effects on word duration because I excluded all trials with pauses longer than 200 ms in one set of models. In a second set of models I included trials with pauses longer than 200 ms but all critical effects remained significant. Furthermore, there was no interaction between the presence of a pause longer than 200 ms and the effect of Word2 frequency or REVS on Word1 duration.

Lastly, the current results could have implications about how speakers are able to achieve fluency. According to Griffin (2003), speakers achieve fluent delivery by estimating word duration. My results suggest that speakers achieve fluent delivery by adjusting Word1 duration. Therefore, a very interesting follow-up to this series of studies would be to understand how these two mechanisms interact when speakers try to achieve fluent delivery.

Were Speakers Planning Serially?

The logic of the experimental manipulation is based on the assumption that speakers plan serially. This means that speakers first look at the left object and then shift their eye gaze to the right object when they have finished phonological planning of the first word (e.g., Meyer & van der Meulen, 2000; Meyer et al., 1998, 2003, 2007). According to this logic the REVS measure should index the onset of right object processing.

However, Malpass and Meyer (2010) found in a dual-picture naming task that the duration of looks to the left object were longer when the picture on the right object was easy to name. This occurred because speakers processed the right object parafoveally, while fixating the left object. This finding suggests that in a dual-picture naming task it is possible to process the two picture names in parallel (at least to a certain degree). Therefore, the Malpass and Meyer's (2010) findings could suggest that the REVS measure did not necessarily index the onset of right object processing. Speakers could have substantially processed the second object prior to fixating it.

The influence of parallel processing on Word1 duration however should have been minimal. In Experiments 3 and 4 there was an interaction between REVS and Word2 frequency when predicting Word1 duration. This suggests that the effect of REVS measure on Word1 duration varied according to Word2 planning. The speakers most probably looked at the left object and then the right object. If the looks to the right object occurred prior to speech onset, the speakers could modulate Word1 duration according to their estimate of Word2 planning difficulty. However, looks that occurred after speech onset did not help in estimating Word2 planning difficulty because Word1 duration did not vary according to Word2 frequency. Under a parallel processing account, knowledge of Word2 planning difficulty prior to fixating the object should have been plausible even when looks to the right object occurred after speech onset. Therefore, Word1 duration should have been modulated according to Word2 frequency independent of REVS.

It is also possible that under a parallel processing account, REVS should have not predicted speech onset time. If looks to the left object provided information about right object processing, then the need to fixate the right object to name it should have been minimized. Nevertheless, REVS was a significant predictor of speech onset time in all experiments with eyetracking. Early looks to the right object led to long speech onset time. Therefore, fixating the right object contributed to an increase in overall processing time, which is not necessarily compatible with a parallel processing account. Furthermore, considerable parallel processing should have rendered REVS a non-significant predictor of Word1 duration.

In conclusion, speakers most probably planned incrementally. However, even if there were some parallel processing of the two objects, Word1 duration variation should be mostly attributed to Word2 planning rather than parallel Word1 and Word2 planning. However, it still remains an open question whether parallel processing of two objects in a dual-naming task could also lead to Word1 duration variation.

Implications about Theories of Planning and Word Duration Variation

The current findings support the proposal that planning can affect word duration. Previous research has found evidence that planning facilitation can lead to word duration variation. (Arnold & Kahn, in press, Arnold & Watson, under review; Balota, Boland & Shield, 1989; Shields & Balota, 1991). The basic reasoning of how facilitation can affect word duration is via speeding up the rate of processing. For example repeated words in the discourse may be reduced because they are facilitated at all levels of representation, semantic, phonological and phonetic/articulatory. According to Balota et al. (1989) the effect of facilitation can be implemented in a connectionist model (e.g., Dell, 1986) by allowing the rate of processing to influence the strength of articulatory movements. Faster rate of processing should lead to faster implementation of articulatory movements and therefore reduced word duration.

There are multiple ways that planning can affect word duration in utterances. For example previous mention can affect word duration but also planning information that is upcoming in the discourse can have an effect (e.g., Christodoulou, 2009; Ferreira &

Swets, 2002). The results presented here suggest that both the timing and duration of planning can affect word duration.

The current results do not support previous research that does not suggest a link between planning and word duration (Damian, 2003; Ferreira, 1991, 2007). For example according to Ferreira (1991, 1993, 2007) word duration should be only determined by the prosodic representation of the utterance. The discrepancy in the results on word duration when manipulating planning with the current results could be attributed to the tasks used. Ferreira (1991) used a prepared speech task, which should minimize the effect of planning on word duration. Damian (2003) used a word interference paradigm, however this task induces a different type of planning than serial planning that should be prevalent in everyday speech (e.g. Dell, Burger, & Svec, 1997).

The Relationship between the Timing of Planning and Word Duration

Clark & Fox Tree (2002) and also Fox Tree & Clark (1997) argue that long word duration can function as a signal to the listener of an upcoming delay. According to this account, word duration variation has primarily a communicative intention. In other words, speakers might use long word duration to heighten the listeners' attention to an upcoming delay.

I cannot completely exclude this account, based on the Schober & Brennan (2003) argument that speakers can design their utterances for a generic addressee even if they are not directly addressing a specific partner. For example, even if a speaker is just recording their voice without addressing a conversational partner, his or her speech could still be influenced by the general tendency to design the utterance for some generic addressee.

However, this kind of a process arguably does not drive the frequency effect on the first word duration. The primary reason is that word variation occurred also when pausing was not disfluent (i.e. less than 200 ms). Therefore, the planning difficulty encountered by the speakers in the current experiments was probably not strong enough to justify the signaling of an upcoming delay.

Previous research suggests that speakers can set a response criterion (Kello et al., 2000; Lupker, Brown, & Colombo, 1997; Meyer et al., 2003; Meyer et al., 2007). This criterion can be set for all trials in an experimental block to aid the speaker with the decision of when to start speaking. By setting the criterion once the speaker does not have to update it on a trial-by-trial basis. Therefore he or she is always prone to start speaking at a specific time point for all trials.

A possible explanation of why word duration varied according to the second word frequency is that speech onset time occurred prior to when speakers were ready to start speaking. Therefore, speakers might have developed a strategy where they always started speaking knowing they would have to use a long first word duration. Because low frequency words require prolonged planning time, the first word duration was longer when followed by a low as opposed to a high frequency word.

According to this account, the effect of REVS on word duration can be interpreted as the effect of variation in the ability to meet that deadline. When the speaker could comfortably meet the deadline, he or she could have sped up the rate of processing Word1 to start planning Word2. However, when the speaker had trouble meeting the deadline, the speaker could have slowed down their rate of processing and as a result delayed planning of Word2.

This interpretation can also accommodate the interaction between REVS and Word1 duration. When speakers could comfortably meet the deadline and Word2 was easy to plan, Word1 duration was reduced. Potentially this could occur because the current rate of processing could accommodate the additional processing of Word2. However, when speakers could comfortably meet the deadline and Word2 was difficult to plan Word1 duration was not reduced. This potentially could occur because the current rate of processing could not easily accommodate the additional processing of Word2.

One problem with the deadline account is that when the experimental manipulation is blocked, there should be a difference in speech onset time according to experimental block. This result occurred in Experiment 4 but not in Experiment 3. Furthermore, according to this account, REVS should have interacted with experiment block in Experiment 4. However, this effect did not occur.

According to a different account, the effect of REVS on Word1 duration could be attributed to variation in the degree of incremental planning. On this account, instead of a deadline to begin speaking, participants could have adopted a processing deadline (e.g., Meyer et al., 2007). Previous research suggests that in the dual-picture naming task speakers usually shift planning to upcoming material when they have completed planning of a minimum amount of phonological information of Word1 (Meyer & van der Meulen, 2000; Meyer et al., 1998; Meyer et al., 2003). For example, Meyer et al. (2007) suggest that this minimum amount could be equal to a single syllable. Therefore, when trials allowed faster completion of phonological Word1 information, they could easily shift to planning Word2.

This account is similar to the previous account in that the effect of early REVS on word duration is not motivated by the difficulty of planning Word2. Rather, it is motivated by random fluctuation around a predetermined deadline. However, the difficulty of planning Word2 modulates the effect of REVS on Word1 duration according to the fluctuation around the deadline.

One problem with this account is that in Experiment 3 REVS did not interact with the Word2 timing manipulation. However, this could potentially be attributed to the already high degree of incremental planning in the experiment. Therefore this second account receives relatively more support compared to the first account.

According to the last account of why REVS influences Word1, the timing of retrieving Word2 information drives the effect. REVS could be indexing variation in the ability to plan together Word1 and Word2 (i.e. coordinate their planning). The task demands required the participants to try not to pause between the bare nouns. During trials where the speaker was able to look at the right object relatively early to speech onset, he or she could have been more prone to meet task demands of not pausing by trying to preplan the two words. This behavior could also have accounted for the interaction between REVS and Word2 frequency. When the speakers looked at the right object early relative to the onset of speech and Word2 frequency was high, Word1 duration was reduced because planning the two words together was easy. However, when Word2 duration was low frequency, planning the two words together was difficult. When looks to the right object occurred late relative to the onset of speech, the ability to plan both words was overall reduced. Therefore Word1 duration was overall reduced.

Future Directions and Conclusions

Follow up studies could try to tease apart the different accounts of how the timing of planning affects Word1 duration. One way of testing whether the degree of coordinating preplanning holds would be to use a dual-picture naming task that allows participants to inspect both objects for a fixed amount of time prior to speech onset. Then vary the amount of preview time. Long preview time should lead to higher ability to preplan the two words therefore Word1 should be reduced.

A way of testing the processing deadline account is to manipulate the number of syllables of Word1 while keeping constant the number of syllables in Word2 (e.g., Meyer et al., 2007). The duration of Word1 should be unaffected by REVS and Word2 difficulty.

Lastly, it would be interesting to explore the idea that speakers are using long word duration to signal to an upcoming delay. I could use a variant of the two-picture naming task where the speaker needs to describe the two objects for a listener to identify on his or her screen. In one condition the time of identification by the listener would be critical for the successful completion of the task. For example the instructions could require the listener to identify both object prior to a certain amount of time. The critical question would be whether the degree of Word1 duration variation would be modulated by the time pressure manipulation and whether it would interact with planning difficulty of Word2.

In conclusion, word duration is related to planning. Speakers used long word durations when the upcoming word was difficult to plan and when the timing of planning the upcoming word was delayed. Both of those findings suggest that longer processing time leads to long word duration.

Appendix A

Items in all four experiments

Item	Freq	Experiment 1 Left	Experiment 1 Right	Experiment 2 Left	Experiment 2 Right	Experiment 3 & 4 Left	Experiment 3 & 4 Right
1	H	skunk	hand	-	-	-	-
1	L	skunk	deer	-	-	-	-
2	H	peach	door	clown	-	-	-
2	L	peach	pants	clown	-	-	-
3	H	axe	dog	-	-	-	-
3	L	axe	spoon	-	-	-	-
4	H	pear	foot	-	-	-	-
4	L	pear	kite	-	-	-	-
5	H	sock	book	-	-	-	-
5	L	sock	sled	-	-	-	-
6	H	flute	car	-	-	-	-
6	L	flute	vest	-	-	-	-
7	H	comb	eye	-	-	-	-
7	L	comb	grapes	-	-	-	-
8	H	glove	heart	-	-	-	-
8	L	glove	broom	-	-	-	-
9	H	harp	bed	-	-	-	-
9	L	harp	wrench	-	-	-	-
10	H	goat	ball	-	-	-	-
10	L	goat	ant	-	-	-	-
11	H	snail	house	-	-	-	-
11	L	snail	owl	-	-	-	-
12	H	stool	tree	-	-	-	-
12	L	stool	swan	-	-	-	-
13	H	zebra	jacket	-	wagon	-	-
13	L	zebra	snowman	-	-	-	-
14	H	ashtray	football	spider	-	-	-
14	L	ashtray	beetle	spider	whistle	-	-
15	H	toaster	chicken	-	-	-	-
15	L	toaster	giraffe	-	-	-	-
16	H	camel	flower	-	-	-	-
16	L	camel	doorknob	-	-	-	-
17	H	raccoon	glasses	-	-	-	-
17	L	raccoon	scissors	-	-	-	-
18	H	hanger	window	-	-	-	-
18	L	hanger	windmill	-	-	-	-
19	H	lettuce	bottle	-	-	monkey	-
19	L	lettuce	lobster	-	-	monkey	-

		Experiment 1		Experiment 2		Experiment 3 & 4	
Item	Freq	Left	Right	Left	Right	Left	Right
20	H	penguin	finger	-	-	-	-
20	L	penguin	ostrich	-	-	-	-
21	H	pumpkin	table	-	-	-	-
21	L	pumpkin	thimble	-	-	-	cannon
22	H	paintbrush	guitar	-	-	-	-
22	L	paintbrush	pliers	-	-	-	-
23	H	donkey	pencil	-	-	-	-
23	L	donkey	squirrel	-	-	-	-
24	H	leopard	mountain	carrot	-	-	-
24	L	leopard	mitten	carrot	-	-	-

Note. Freq refers to the frequency category of the right picture within each item. A dash suggests that the picture did not change between experiments. Experiments 3 and 4 had the same items.

Appendix B

Model specifications and model criticisms for Experiment 1

Table B1

Experiment 1 Word1 duration model random effect structure

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0032	0.0569
Sub	W2freq	0.0002	0.0129
Item	Intercept	0.0028	0.0525
Item	W2freq	0.0006	0.0246
Residual		0.0055	0.0742

Table B2

Experiment 1 Word1 duration model fixed effects partial correlations

	Intercept	Syllables	Pause	Trial#	W1onset
Syllables	-0.03				
Pause	-0.016	0.049			
Trial#	0.004	-0.031	0.017		
W1onset	0.001	-0.002	-0.062	-0.027	
W2freq	0.008	0.004	-0.046	-0.012	0.014

Note. Trial# refers to trial order. W1onset refers to the Word1 initial phoneme category.

Table B3

Experiment 1 Speech onset time model random effect structure

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0059	0.0768
Item	Intercept	0.0009	0.0298
Item	W2freq	0.0013	0.0363
Residual		0.0046	0.0680

Table B4

Experiment 1 Speech onset time model fixed effects partial correlations

	Intercept	Pause	Trial#	W2H
Pause	-0.015			
Trial#	0.003	0.007		
W2H	-0.007	-0.019	0.052	
W2freq	0.007	-0.021	-0.001	0.036

Note. Trial# refers to trial order. W2H refers to the Word2 codability.

Table B5

Experiment 1 fixed effects for word duration model with pauses longer than 200 ms

	b	*SE*	*t*	*P*
Intercept	2.674	0.017	157.440	0.0000
Syllables	0.075	0.027	2.810	0.0077
Pause	0.022	0.006	3.700	0.0017
Trial#	0.000	0.000	1.190	0.2657
Sex	0.035	0.025	1.440	0.1630
W1onset	0.024	0.011	2.140	0.0287
W1offset	0.012	0.008	1.480	0.1409
W2freq	0.042	0.011	3.650	0.0002
X200	0.026	0.018	1.430	0.0697
X200*W2freq	-0.026	0.028	-0.930	0.3368

Note. Trial# refers to trial order. W1onset refers to Word1 initial phoneme category. Word1offset refers to Word1 final phoneme category. X200 refers to the presence of a pause longer than 200 ms

Table B6

Experiment 1 random effect structure for word duration model with pauses longer than 200 ms

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0029	0.0536
Sub	W2frequency	0.0000	0.0000
Item	Intercept	0.0032	0.0563
Item	W2frequency	0.0013	0.0359
Residual		0.0058	0.0764

Table B7

Experiment 1 fixed effects partial correlations for word duration model with pauses longer than 200 ms

	Intercept	Syllable	Pause	Trial#	Sex	W1o	W2frq	X200
Syllables	-0.022							
Pause	-0.001	0.018						
Trial#	0.002	-0.031	0.016					
Sex	-0.029	0.002	0.014	-0.004				
W1onset	0.016	-0.038	-0.03	-0.019	0.001			
W1offset	0.016	-0.225	0.048	0.021	0	0.187		
W2freq	-0.001	0.007	-0.038	-0.003	0.011	0.015	0	
X200	0.013	-0.031	-0.619	-0.063	-0.032	-0.001	-0.014	-0.121
X200*W2freq	-0.053	0.009	-0.032	0.031	0.056	0.007	-0.003	0.05

Note. Trial# refers to trial order. W1onset refers to Word1 initial phoneme category. W1offset refers to Word1 final phoneme category. X200 refers to the presence of a pause longer than 200 ms

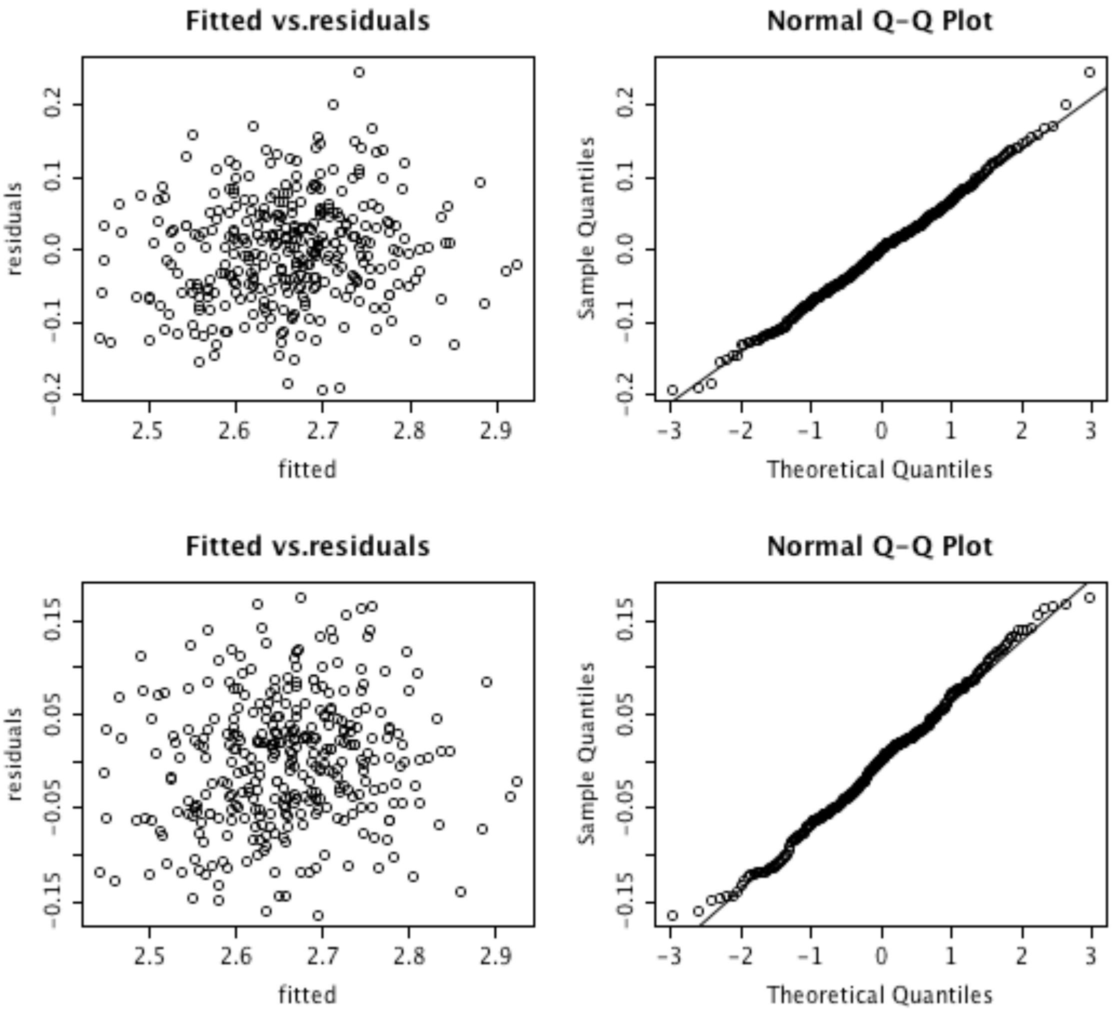


Figure B1. The top row represents model criticism graphs for the Word1 duration model in Experiment 1 before outlier exclusions. The second row represents model criticism graphs after outlier exclusions (n=3).

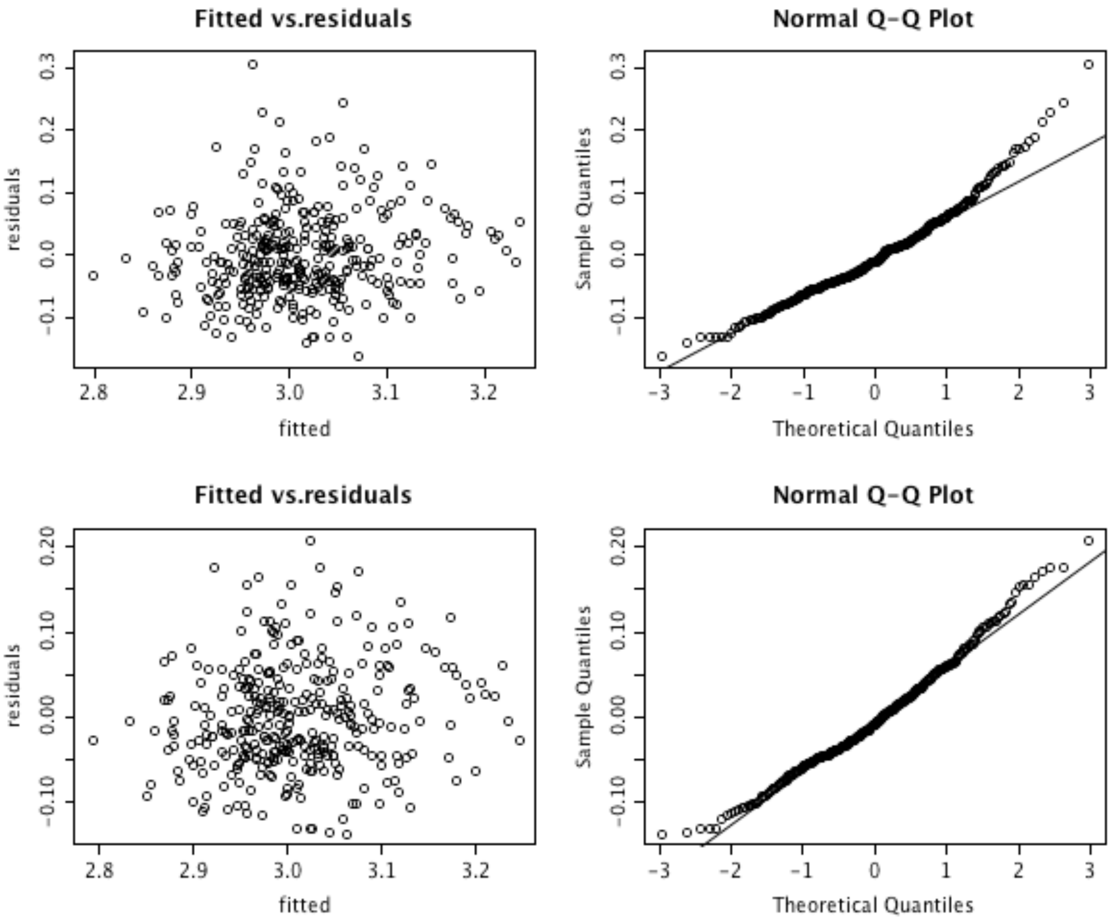


Figure B2. The top row represents model criticism graphs for the speech onset time model in Experiment 1 before outlier exclusions. The second row represents model criticism graphs after outlier exclusions (*n*=2).

Appendix C

Model specifications and model criticisms for Experiment 2

Table C1

Experiment 2 Word1 duration model random effect structure

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0074	0.0863
Sub	W2freq	0.0000	0.0000
Item	Intercept	0.0034	0.0586
Item	W2freq	0.0005	0.0228
Residual		0.0047	0.0682

Table C2

Experiment 2 Word1 duration fixed effect partial correlations

	Intercept	Pause	Syllables	Trial#	W1onset	Sdis	W2freq
Pause	-0.018						
Syllables	-0.006	0.035					
Trial#	-0.003	-0.033	0.005				
W1onset	-0.033	-0.004	-0.098	0.001			
Sdis	-0.025	-0.041	0.04	0.078	-0.147		
W2freq	0.059	-0.136	-0.024	0	0.031	-0.074	
REVS	-0.006	-0.159	0.019	-0.035	-0.058	0.059	0.155
W2freq* REVS	0.029	-0.116	-0.019	0.013	0.004	0.092	0.064

Note. Trial# refers to trial order. W1onset refers to Word1 initial phoneme category. Sdis refers to the disfluency rate by subjects.

Table C3

Experiment 2 Speech onset time model random effects structure

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0032	0.0566
Sub	W2freq	0.0000	0.0000
Item	Intercept	0.0019	0.0433
Item	W2freq	0.0000	0.0000
Residual		0.0069	0.0831

Table C4

Experiment 2 Speech onset time fixed effects partial correlations

	Intercept	Trial#	Serr	W1H	W2freq
Trial#	-0.01				
Serr	-0.034	0.022			
W1H	-0.026	-0.041	0.023		
W2freq	-0.002	0.013	0.26	0.035	
REVS	-0.023	-0.029	0.034	-0.012	0.175
W2freq*REVS	0.049	0.001	-0.08	0.027	0.019

Note. Trial# refers to trial order. Serr refers to the error rate by subjects. W1H refers to the Word1 codability.

Table C5

Experiment 2 fixed effects for word duration model with pauses longer than 200 ms

	b	*SE*	*t*	*p*
Intercept	2.62000	0.02196	0.00000	0.00000
Syllables	0.01148	0.00611	0.06140	0.06140
Pause	0.04854	0.02807	0.08470	0.08470
Trial#	0.00103	0.00029	0.00030	0.00030
W1onset	0.02090	0.01208	0.08440	0.08440
W2onset	-0.00730	0.00459	0.11300	0.11300
Idis	0.20160	0.14080	0.15320	0.15320
W2H	0.02940	0.01751	0.09410	0.09410
W2freq	0.02746	0.00849	0.00130	0.00130
REVS	0.00009	0.00002	0.00000	0.00000
X200	-0.00113	0.01776	0.94940	0.94940
W1off*W2on	0.00658	0.00442	0.13740	0.13740
W2freq*REVS	-0.00003	0.00003	0.37670	0.37670
X200*W2freq	0.03051	0.02606	0.24260	0.24260
X200*REVS	-0.00001	0.00005	0.88300	0.88300
X200*W2freq*REVS	0.00006	0.00009	0.50820	0.50820

Note. Trial# refers to trial order. W1onset and W2onset refer to the Word1 and Word2 initial phoneme category. Idis refers to the disfluency rate by Items. W2H refers to the Word2 codability. X200 refers to the presence of a pause longer than 200 ms.

Table C6

Experiment 2 random effects for word duration

model with pauses longer than 200 ms

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0068	0.0826
Sub	W2feq	0.0000	0.0000
Item	Intercept	0.0034	0.0580
Item	W2freq	0.0001	0.0083
Residual		0.0047	0.0687

Table C7

Experiment 2 fixed effects partial correlations for the word duration model with pauses longer than 200 ms

	Intercept	Syllables	Pause	Trial	W1onset	W2onset	W2H
Intercept	-0.004						
Syllables	0	0.001					
Pause	-0.001	-0.03	-0.002				
Trial	-0.019	-0.002	-0.077	0.014			
W1onset	0.007	0.117	-0.185	-0.003	-0.06		
W2onset	0.011	-0.056	-0.014	0.013	-0.102	0.1600	
Items dis	-0.025	-0.023	-0.067	-0.014	0.037	-0.2850	
W2H	0.019	-0.164	-0.05	0.032	0.019	0.1180	-0.009
W2freq	0.002	-0.119	0.034	-0.028	-0.063	-0.0570	0.021
REVS	0.008	-0.615	0.01	0.038	-0.013	-0.0780	0.013
X200	-0.046	-0.011	-0.012	0.029	0.049	-0.2290	0.408
W1off*W2on	0.03	-0.094	-0.022	0.033	-0.002	0.033	-0.07
W2freq*REVS	-0.014	-0.018	-0.035	0.039	0.033	0.033	0.062
X200*W2freq	-0.035	0.124	0.013	-0.046	0.003	-0.0080	0.11
X200*REVS	0.004	-0.003	-0.001	0.058	-0.005	-0.0440	-0.088

Table C7 Continued

	W2freq	REVS	X200	W1off* W2on	W2freq* REVS	X200* W2freq	X200* REVS
Intercept							
Syllables							
Pause							
Trial							
W1onset							
W2onset							
Idis							
W2H							
W2freq	0.192						
REVS	0.035	0.058					
X200	-0.13	-0.043	0.08				
W1off*W2off	0.031	-0.17	-0.018	-0.096			
W2freq*REVS	0.074	-0.012	-0.127	-0.001	-0.18		
X200*W2freq	-0.092	0.004	-0.321	0.084	-0.144	0.112	
X200*REVS	-0.147	-0.14	0.078	-0.042	0.045	-0.298	-0.408

Note. The notation is the same as table C5.

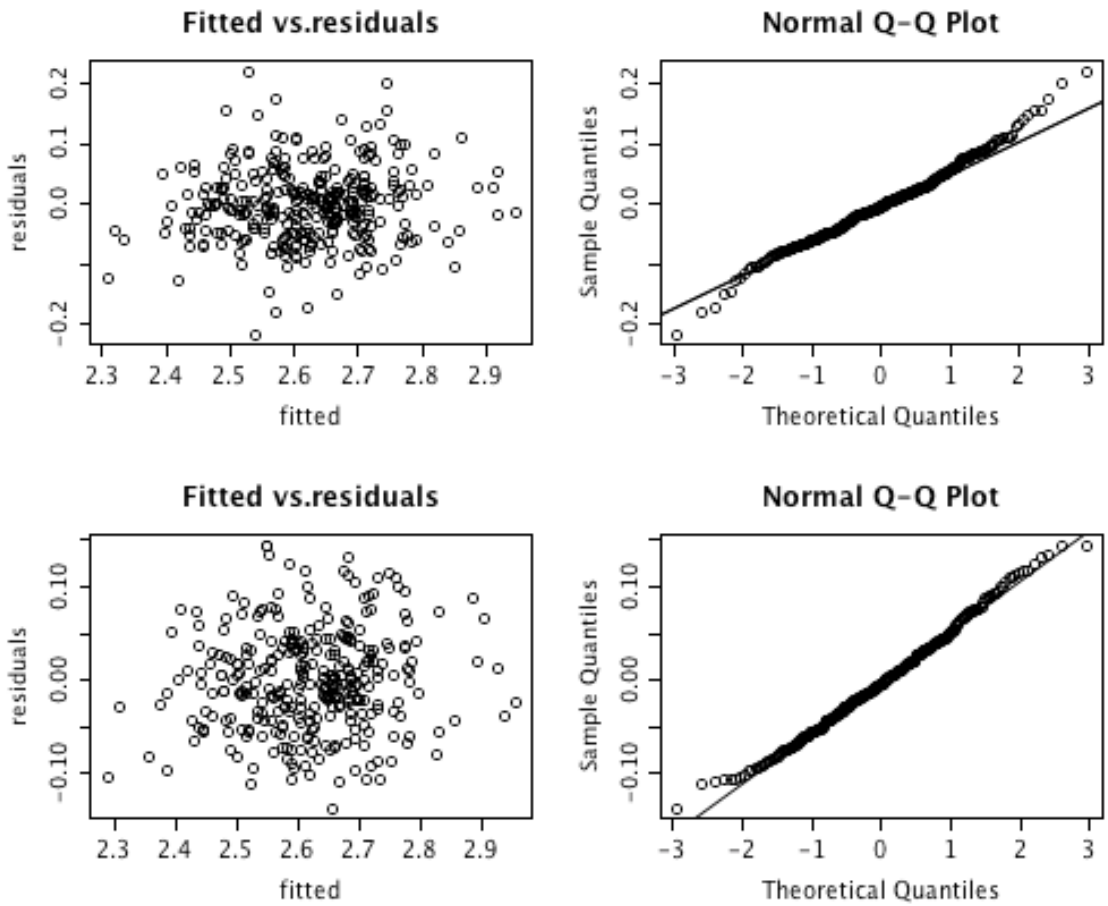


Figure C1. The top row represents model criticism graphs for the Word1 duration model in Experiment 2 before outlier exclusions. The second row represents model criticism graphs after outlier exclusions (n=8).

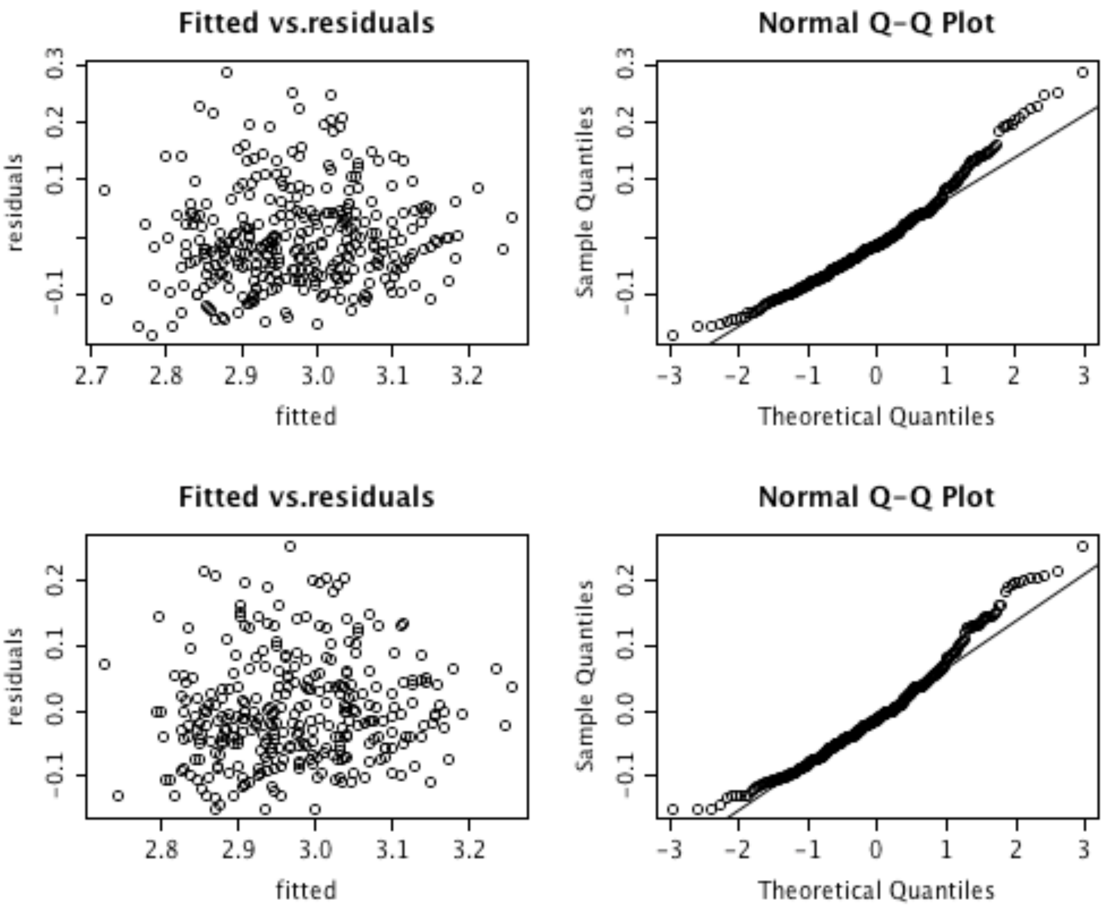


Figure C2. The top row represents model criticism graphs for the speech onset time model in Experiment 2 before outlier exclusions. The second row represents model criticism graphs after outlier exclusions (n=8).

Appendix D

Model specifications and model criticisms for Experiment 3

Table D1

Experiment 3 Word1 duration model random effects structure

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0036	0.0601
Sub	W2freq	0.0000	0.0000
Sub	Instructions	0.0024	0.0486
Item	Intercept	0.0015	0.0386
Item	W2freq	0.0007	0.0263
Item	Instructions	0.0000	0.0023
Residual		0.0039	0.0628

Table D2

Experiment 3 Word1 duration fixed effects partial correlations

	Intercept	Pause	Syllables	First	Trial#	W1onset
Pause	-0.01					
Syllables	0.002	0.055				
First	0.006	0.066	0.011			
Trial#	0.001	-0.018	0.007	-0.007		
W1onset	-0.02	0.032	-0.113	0	0.016	
W1offset	0.039	0.037	-0.096	0.001	0.007	0.249
W2freq	0.128	-0.053	-0.012	0.017	-0.035	-0.022
REVS	-0.003	-0.068	-0.009	0.025	0.009	-0.057
Instructions	0.292	0.02	0.003	-0.008	-0.009	0.015
W2freq*REVS	0.009	0.037	0.005	-0.038	-0.071	0.029

Table D2 Continued

	W1offset	W2freq	REV
Pause			
Syllables			
First			
Trial#			
W1onset			
W1offset			
W2freq	0.012		
REV	0.007	0.043	
Instructions	0	-0.014	-0.146
W2freq*REV	-0.011	0.003	-0.08

Note. First refers to the experimental block order. Trial# refers to trial order. W1onset and W1offset refers to Word1 initial and final phoneme category.

Table D3

Experiment 3 Speech onset time model random effects structure

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0029	0.0537
Sub	W2freq	0.0002	0.0125
Sub	instructions	0.0008	0.0279
Item	Intercept	0.0012	0.0340
Item	W2freq	0.0001	0.0119
Item	instructions	0.0000	0.0000
Residual		0.0045	0.0673

Table D4

Experiment 3 Speech onset time fixed effects partial violations

	Inter	Versn	First	Ierr	W1H	W2H	W2frq	REV	instr
Version	0.032								
First	0.002	0.014							
W1fam	0.033	0	-0.004						
W1H	0.006	0.001	0.004	0.128					
W2H	0.001	0.008	0.003	0.138	0.051				
W2freq	0.001	-0.003	0.019	0.021	0.011	0.132			
REV	-0.011	-0.05	0.035	-0.008	-0.005	-0.04	0.054		
Instructions	-0.005	0.023	-0.008	-0.003	0.004	-0.028	-0.077	-0.198	
W2fq*REV	0.013	-0.004	-0.047	-0.004	-0.002	0.046	0.004	-0.074	0.019
Instr*REV	-0.054	0.009	-0.016	-0.025	0.001	-0.067	-0.004	-0.014	0.028

Note. Version refers to the combination of experimental block and experimental list. First refers to the experimental block order. W1H and W2H refers to Word1 and Word2 codability. Instructions refers to the timing manipulation.

Table D5

Experiment 3 fixed effects for word duration model with pauses longer than 200 ms

	b	*SE*	*t*	*P*
Intercept	2.68300	0.01298	206.67000	0.0000
Pause	0.01043	0.00299	3.49000	0.0005
Syllables	0.03111	0.01848	1.68000	0.0926
First	-0.02745	0.01073	-2.56000	0.0107
Trial#	0.00064	0.00022	2.86000	0.0044
W1onset	0.03001	0.00754	3.98000	0.0001
W1offset	0.01509	0.00516	2.92000	0.0036
W1freq	0.01717	0.00673	2.55000	0.0109
REVS	0.00005	0.00001	3.93000	0.0001
X200	0.00198	0.00907	0.22000	0.8273
Instructions	-0.00194	0.01140	-0.17000	0.8647
W2freq*REV	-0.00004	0.00002	-2.21000	0.0271
W2freq*X200	0.01487	0.01486	1.00000	0.3172
REVS*X200	-0.00003	0.00002	-1.23000	0.2200
W2freq*REVS*X200	-0.00002	0.00005	-0.34000	0.7342

Note. First refers to the experimental block order. Trial# refers to the trial order. W1onset and W1offest refer to Word1 onset and offset phoneme category. X200 refers to the presence of a pause longer than 200 ms.

Table D6

Experiment 3 fixed effects partial correlations for word duration model with pauses longer than 200 ms

	Intercept	Pause	Syls	First	Trial#	W1on	W1off
Pause	0.003						
Syllables	-0.006	0.046					
First	0.009	0.046	0.002				
Trial#	0.001	0.014	0.007	0.001			
W1onset	-0.021	0.019	-0.123	0.001	0.015		
W1offset	0.024	0.024	-0.095	0.003	0.004	0.262	
W1frequency	0.001	-0.055	-0.008	0.021	-0.033	-0.007	-0.003
REVS	0.007	-0.083	-0.006	0.019	-0.024	-0.046	-0.005
X200	0.004	-0.422	-0.013	-0.001	-0.033	-0.021	0.028
Instructions	-0.002	0.03	0.002	-0.022	-0.001	0.011	0.003
W2freq*REV	0.01	0.056	0.007	-0.032	-0.053	0.004	-0.008
W2freq*X200	-0.025	-0.019	0.002	-0.009	0.03	0.018	-0.015
REVS*X200	-0.022	0.042	0	0	-0.023	0.007	-0.001
W2frequency *REVS*X200	0.004	0.01	0.001	0.02	0.053	-0.01	0.005

Table D6 Continued

	W1freq	REVS	X200	Instr	W2freq *REV	W2freq *X200	REVS *X200
Pause							
Syllables							
First							
Trial#							
W1onset							
W1offset							
W1frequency							
REVS	0.052						
X200	-0.063	-0.02					
Instructions	-0.016	-0.126	0.054				
W2freq*REV	0.005	-0.053	-0.031	0.014			
W2freq*X200	0.032	0.015	-0.246	-0.004	-0.11		
Rev*X200	0.001	-0.119	-0.17	0.024	-0.022	0.073	
W2frequency *REVS*X200	-0.077	-0.019	0.106	0.009	-0.132	-0.244	-0.125

Note. The notation is the same as table D5

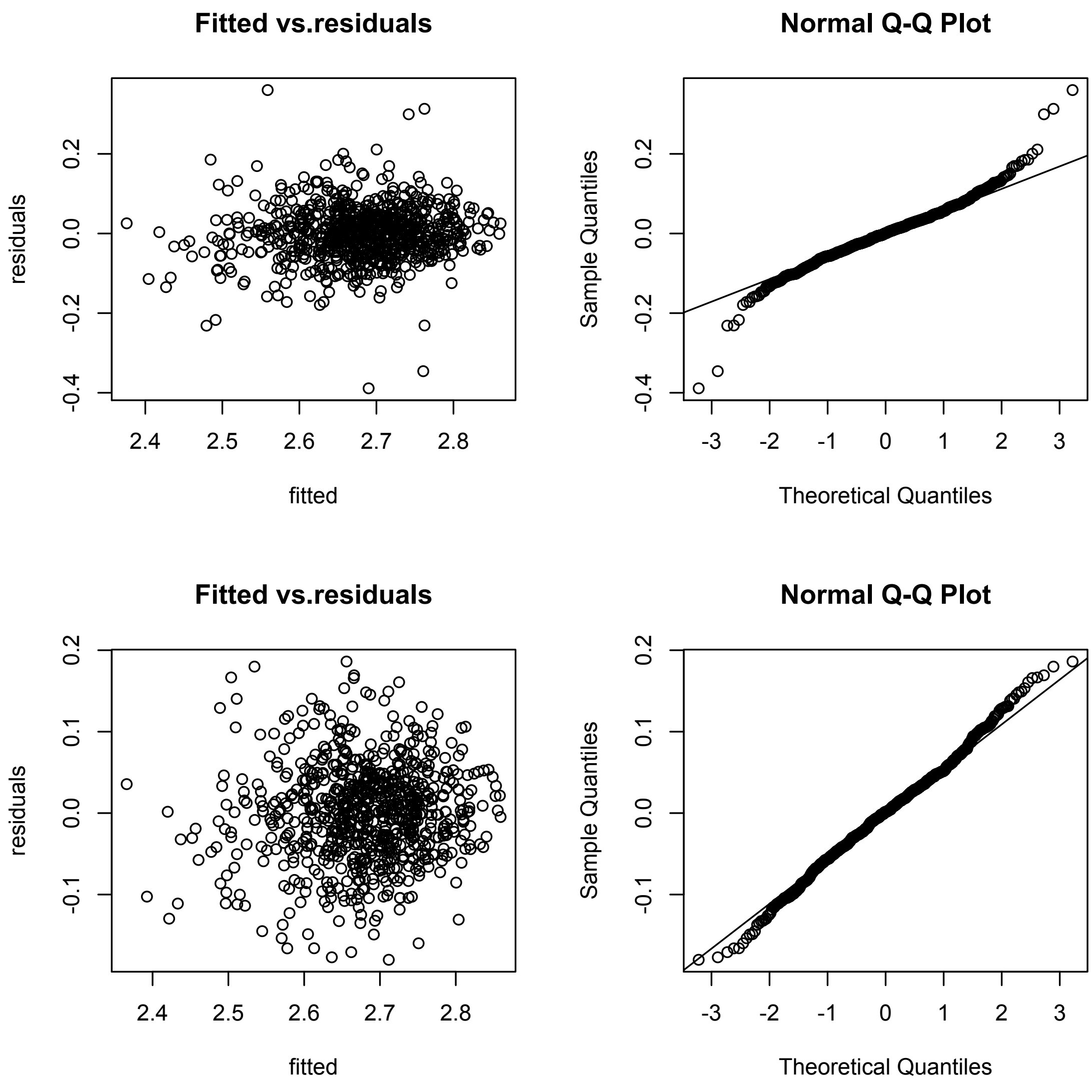


Figure D1. The top row represents model criticism graphs for the Word1 duration model in Experiment 3 before outlier exclusions. The second row represents model criticism graphs after outlier exclusions (n=10).

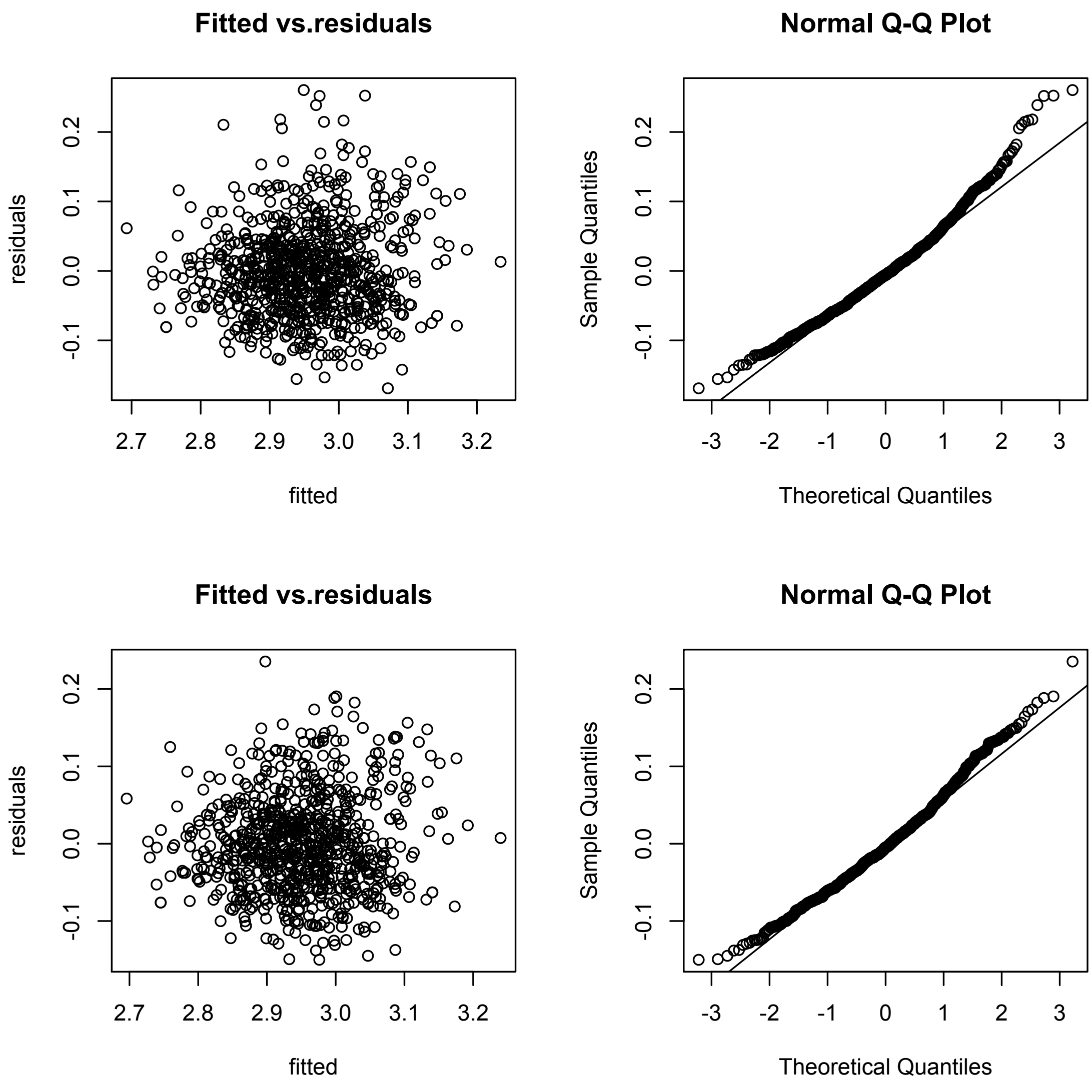


Figure D2. The top row represents model criticism graphs for the speech onset time model in Experiment 3 before outlier exclusions. The second row represents model criticism graphs after outlier exclusions (n=8).

Appendix E

Model specifications and model criticisms for Experiment 4

Table E1

Experiment 4 Word1 duration model random effects structure

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0035	0.0591
Sub	W2freq	0.0000	0.0000
Sub	Instructions	0.0009	0.0300
Item	Intercept	0.0015	0.0384
Item	W2freq	0.0011	0.0328
Item	Instructions	0.0000	0.0000
Residual		0.0036	0.0602

Table E2

Experiment 4 Word1 duration fixed effects partial correlations

	Intercept	Pause	First	Syllables	W1onset	W1offset	Instr
Pause	-0.026						
First	0.003	0.026					
Syllables	-0.014	0.071	-0.006				
W1onset	0.005	0.052	0.014	-0.177			
W1offset	0.042	0.084	0.002	-0.117	0.34		
Instr	-0.006	0.074	0.01	0.005	0.006	0.008	
W2freq	0.006	-0.029	0.003	-0.019	-0.001	0.001	0.006
REVS	-0.011	-0.086	-0.044	0.02	-0.072	-0.05	-0.078
W2freq* Instr	0.001	-0.015	0.007	0.002	-0.006	-0.012	-0.005
W2freq* REVS	0.012	0.032	0.003	0.002	0.015	-0.003	0.041
Pause* Instr	0.002	0.017	0.024	0.003	-0.013	0.003	-0.017
Pause* Syllables	0.033	0.019	0.027	-0.027	0.042	0.003	-0.005
Pause* Ins*RVS	0.011	-0.094	0.016	-0.013	0.01	0.024	-0.116
Pause* W2fq* RVS	0.002	0.004	0.015	-0.009	0.005	0.005	-0.003
Instr* W2frq* RVS	0.003	0.032	0.017	0.005	0.019	0.006	0.008

Table E2 continued

	W2freq	REVS	W2frq *Instr	W2freq *REVS	Pause *Instr	Pause *Syls	Pause *Instr *RS	Pause *W2q *RS
Pause								
First								
Syllables								
W1onset								
W1offset								
Instr								
W2freq								
REVS	0.022							
W2freq*Instr	-0.011	0.088						
W2freq*REVS	0.009	-0.231	-0.16					
Pause*Instr	-0.002	0.033	0.028	-0.023				
Pause*Syllables	0.049	0.003	0.006	-0.036	-0.024			
Pause*Ins* REVS	-0.008	-0.262	-0.027	0.01	-0.011	-0.014		
Pause*W2fq* RVS	-0.093	-0.016	0.03	0.096	0.055	0.029	0.044	
Instr*W2frq* REVS	-0.002	-0.014	-0.026	0.183	-0.076	-0.009	0.007	0.023

Note. First refers to the experimental block order. W1onset and W1offest refer to Word1 onset and offset phoneme category. Instr refers to the timing manipulations.

Table E3

Experiment 4 Speech onset time random effects structure

Groups	Name	Variance	Std.Dev.
Sub	Intercept	0.0016	0.0398
Sub	W2freq	0.0002	0.0126
Sub	Instructions	0.0007	0.0259
Item	Intercept	0.0012	0.0353
Item	W2freq	0.0002	0.0154
Item	Instructions	0.0005	0.0221
Residual		0.0030	0.0548

Table E4

Experiment 4 Speech onset time model fixed effects partial correlations

	Intercept	Pause	First	W1H	Instr
Pause	-0.031				
First	-0.001	0.032			
W1H	0.025	0.012	-0.005		
Instr	0.258	0.041	0.011	0.01	
W2freq	0.111	-0.018	0.013	0.016	0.05
REVS	-0.016	-0.133	0.016	0.012	-0.07
Feed*Instr	-0.115	0.019	0.037	-0.035	0.115
Feed*W2freq	-0.003	0.037	0.001	0.013	-0.003
Feed*REVS	0.01	-0.02	0.077	0.003	0.039
W2Freq*REVS	0.008	0.035	0.004	-0.018	0.058
Feed*W1Freq*REVS	0.002	0.034	-0.012	-0.003	0.077

Table E4 Continued

	W2frq	REVS	Fd*Int	Feed* W2frq	Fd* REVS	W2Frq* REVS
Pause						
First						
W1H						
Instr						
W2freq						
REVS	0.029					
Feed*Instr	-0.031	0.069				
Feed*W2freq	0.016	0.12	0.097			
Feed*REVS	0.114	0.114	-0.015	0.235		
W2Freq* REVS	0.019	-0.314	0.093	0.039	0.034	
Feed* W1Freq* REVS	0.032	0.046	0.19	0.082	-0.168	0.149

Note. First refers to the experimental block order. W1H refers to Word1 codability. Instr refers to the timing manipulations. Feed refers to the feedback per trial.

Table E5

Experiment 4 Word1 duration fixed effect model estimates after including pauses longer than 200 ms

	b	*SE*	*t*	*P*
Intercept	2.6580	0.0146	181.8100	0.0000
Pause	0.0100	0.0027	3.7400	0.0002
First	-0.0075	0.0062	-1.2100	0.2250
Syllalbes	0.0524	0.0193	2.7200	0.0067
W1onset	0.0218	0.0079	2.7400	0.0062
W2offset	0.0176	0.0054	3.2400	0.0013
Instructions	-0.0313	0.0066	-4.7000	0.0000
W2freq	0.0298	0.0069	4.3000	0.0000
REVS	0.0001	0.0000	6.2000	0.0000
Pause*Syllables	0.0127	0.0045	2.7900	0.0054
Instr*W2freq	0.0036	0.0084	0.4300	0.6708
Instr*REVS	0.0001	0.0000	1.8700	0.0616
W2freq*REVS	-0.0001	0.0000	-2.1200	0.0344
Syls*REVS	0.0000	0.0000	-1.7500	0.0808
Syls*Instr	0.0151	0.0094	1.6100	0.1083
First*Instr	0.0960	0.0453	2.1200	0.0345
Pause*Instr*REVS	0.0000	0.0000	-2.1000	0.0360
Pause*W2freq&REV	0.0000	0.0000	-2.3000	0.0219
Pause*Syls*Instr	-0.0209	0.0079	-2.6300	0.0086
First*Instr*REVS	-0.0001	0.0001	-2.1600	0.0308

Note. First refers to the experimental block order. W1onset and W1offest refer to Word1 onset and offset phoneme category. Instructions refers to the timing manipulations.

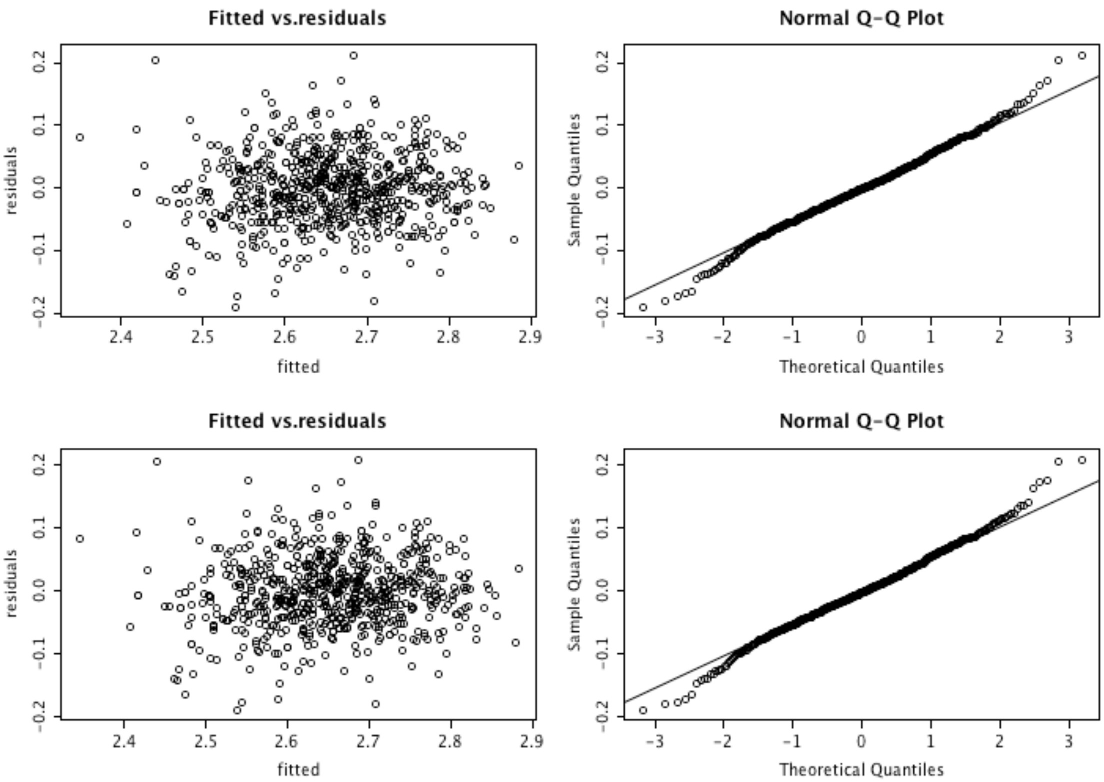


*Figure E1***.** The top row represents model criticism graphs for the Word1 duration model in Experiment 4 before outlier exclusions. The second row represents model criticism graphs after outlier exclusions (*n*=7).

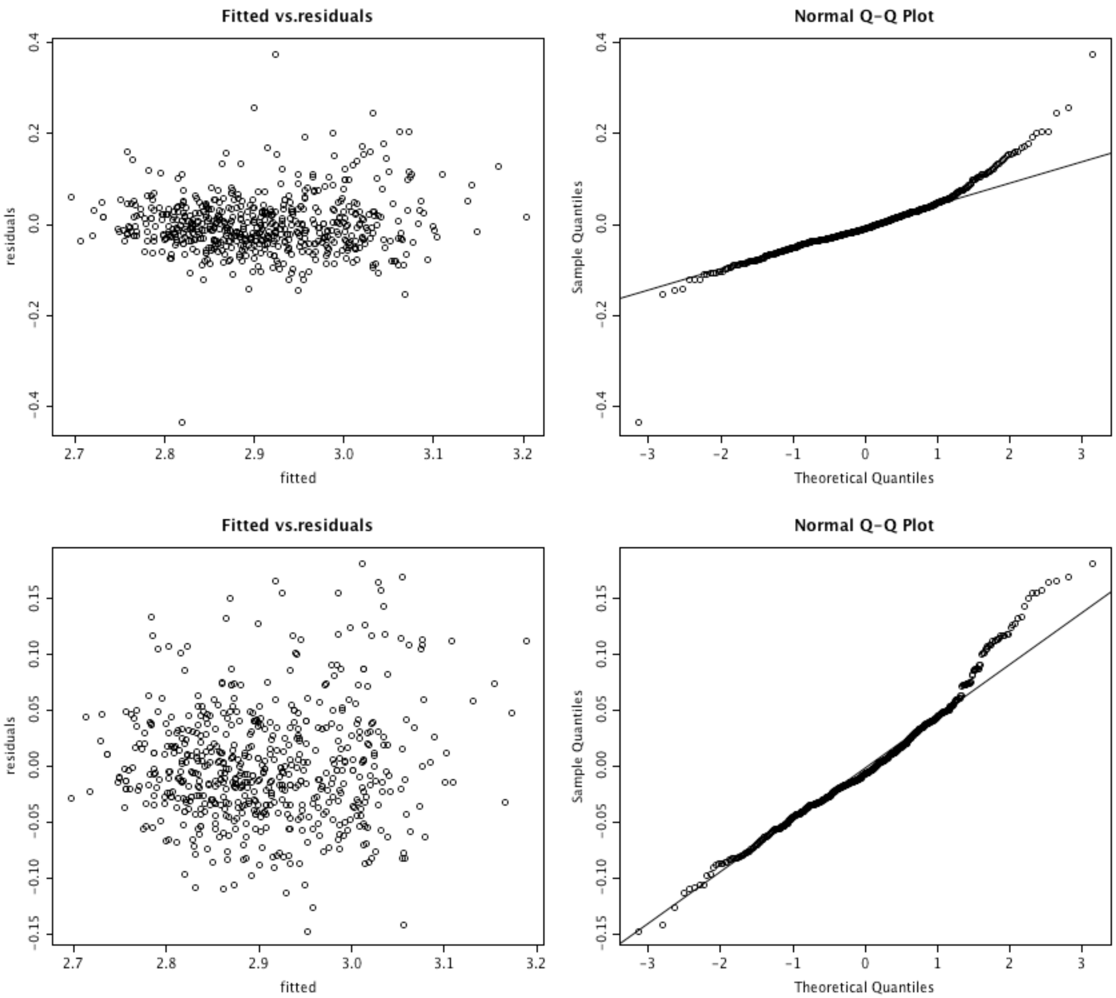


*Figure E2***.** The top row represents model criticism graphs for the speech onset model in Experiment 4 before outlier exclusions. The second row represents model criticism graphs after outlier exclusions (*n*=10)

Appendix F

Processing Sounds

This appendix contains the manual my research assistants and I used to code acoustic data for all the experiments. The process included a first pass autosegmentation of the sounds using either the P2FA aligner or the Prosodylab-Aligner. At a second stage either I or a research assistant hand corrected inaccuracies in boundary alignment produced by the autosegmenation algorithm. The different headers in the document correspond to separate stages in implementing the autosegmentation and then the had correction. In each section I have also included scripts that I developed to automate processing of the sound files. I have also included some screen shots that help illustrate how boundaries were adjusted after the sounds were first autosegmented.

Required software and scripts

1. Praat (http://www.fon.hum.uva.nl/praat/)
2. HTK 3.4 for UNIX (http://htk.eng.cam.ac.uk/) and required subcomponent software
    a. Don't install the latest version (3.4.1) because it won't work!
    b. You will need the latest xcode developer tools to compile. For the 10.4 system you will need xcode 2.5. (http://developer.apple.com/technologies/tools/xcode.html)
3. Python 2.5 or 2.6 (http://python.org/)
4. P2FA (http://www.ling.upenn.edu/phonetics/p2fa/) or Prosodylab Aligner (http://prosodylab.org/tools/aligner/)

5. Optionally: GSU praat tools

   (https://sites.google.com/site/psyvoso/SoftwareDownload)

Create TextGrids

Open and select all sound files in praat. Select Annotate – To TextGrid… This creates empty .TextGrids for all selected files

Create text files

The text files are transcriptions of the recorded utterances and they need to have the same name as the sound file and the corresponding .TextGrid. Because we know the words the speaker is expected to say, we can automate the transcription process as follows: Create an excel file that pairs information of each picture pair with their filename.

The first two columns represent the names of the pairs of left and right objects and the last column their ID. Concatenate the names in the two columns and leave a space in between the two names. Replace the two columns with the new column that contains the two names with the space in between. Then run the excel macro I have created that generates .txt or .lab files named after the ID and containing the two picture names with a space in between. I am pasting here the excel macro.

```
#####################################################################
Sub maketexts()
Dim r As Range, ff As Integer
For Each r In Range("A2", Range("a" & Rows.Count).End(xlUp))
        ff = FreeFile
        Open r.Offset(, 1).Value For Output As #ff
            Print #ff, StrConv(r.Offset(, 0).Value, Unicode)
        Close #ff
Next
End Sub
#####################################################################
```

Run forced aligner

The p2fa forced aligner runs on a single sound file at a time. I created a bash script that iterates through all the files in a directory. I am pasting the script here.

```
#####################################################################

#!/bin/bash

#####################################################################
#
# This script loops through all files in a directory and processes each
# identically named triplet of .wav .txt .TexGrid files with align.py
# The script will work as long as it is stored in the same directory
# with the files needed for the analysis. The align.py script does not
# have to be in the same directory but make sure to provide the path.
# For example, align.py is stored in ~/documents/p2fa on my computer,
# change this path to reflect its location on your computer.
#
# call the script by cding to the folder with the script and files and
# typing $ bash batch_align.script . Then hit enter and sit back and
# relax as all word boundaries are magically identified and coded!
#
# Alex Christodoulou, Psychology UNC Chapel Hill
# 03/08/11
#
# Thanks to Ryan Scotton for help with this script!
#
#####################################################################
```

```
#list all files that end in .wav and store them in a variable
allWavFiles=`ls *.wav`

#create a loop
for wavFile in $allWavFiles
#call align.py and execute on identically named .wav .txt .TextGrid files
do ~/documents/p2fa/align.py ${wavFile} ${wavFile%\.*}.txt
${wavFile%\.*}.TextGrid
done
exit #optional line, will exit either way :)
####################################################################
```

You need to edit the script to point to the directory the align.py file is in. Also make sure that the align.py file is in the p2fa folder with all the downloaded components. Look for the do command and then edit the directory. For example, mine is in ~/documents/p2fa/align.py. Assuming the batch script and all files are in the same directory you can call the script as follows:

1. cd into the folder where the scripts and files are stored
2. type in terminal: bash batch_align.script
3. hit return
4. In the terminal you will see a creation of a series of .tmp files. If the aligner fails for a specific file, you will get a warning on the screen with the name of the file. The script does not crash, it stops iterating when all files have been processed.

You can also run the forced aligner on a single file.

1. cd into the folder with your scripts and data
2. type in terminal: python aligner.py filename.wav filename.txt filename.TextGrid

Inspect aligned files

Open all files with their associated .TextGrids in praat. Make sure that you have installed GSU praat tools in praat. If all files are saved in Praat_Data, start Praat, click Open and select Read sounds and labels. Hit ok or apply and all .wav files with their associated .TextGrids will load into praat. Select the first sound file and click Edit with Labels. When you are done reviewing the file, select File, Next, to go to the next sound and .TextGrid.

The forced alignment algorithm creates two separate tiers for the individual phones and the word forms. Tier 1 codes for phones and Tier 2 for wordforms. When the algorithm detects a pause it tags it with sp, which stands for short pause. If you include in the transcription files the code {ns} before or between words, it will try to identify background noise. Sp and ns are coded in both tiers. In tier 1 each segment is labeled with the corresponding phone. In tier 2 each segment is labeled with the corresponding word in capitals.

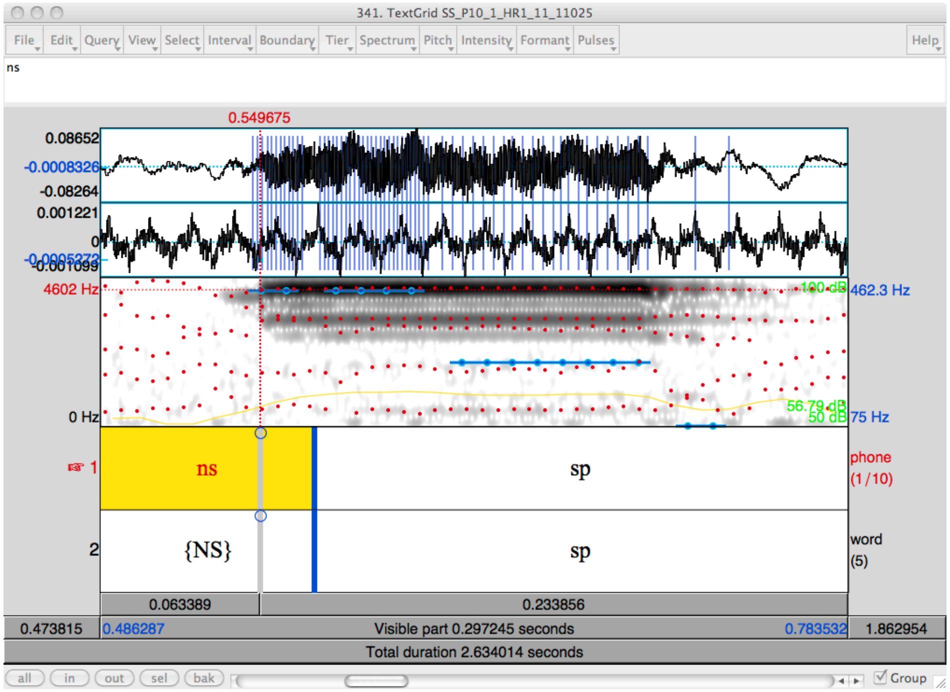
Adjust transcription files

The forced aligner takes the transcription files and tries to identify boundaries no matter what the actual word is. If there is a mismatch between the transcription and the sound, it will still create boundaries according to the transcription file. These files should not be analyzed but if you want to have durations for them, edit the transcription files to reflect the words used.

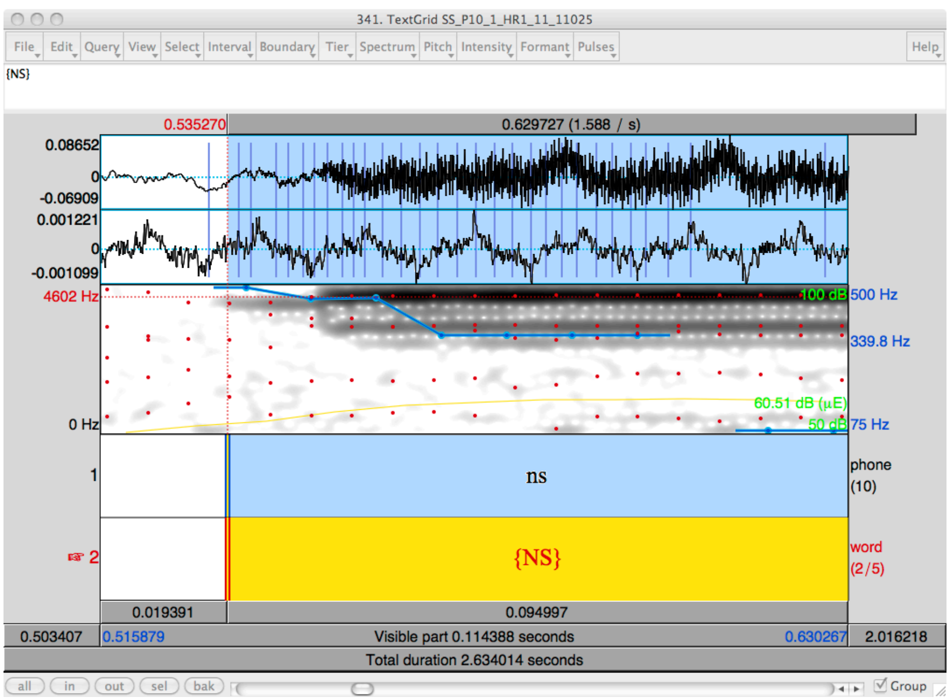
Adjust onset of beep

In most cases the algorithm does not do a good job at detecting the onset of the beep sound at the beginning of each trial. Also, it might detect short pauses between the beep sound and the onset of the first word. Originally in the transcription files I had included {ns} before the two word transcription. The {ns} coding detects noise in the sound file. However, it doesn't always detect the beep sound, it sometimes detects other noises or inaccurately detects the boundaries of the beep. You can either not include the {ns} code in your transcriptions and then manually adjust the onset of the beep sound, or you can adjust the onset of the beep sound manually after it has been automatically detected. The beep sound has a frequency of around 1923 hz which is constant for about 200 ms so it should be fairly easy to identify its onset. However, with this second method you might have to do more editing than just leaving the {ns} code out of the transcription. If you adjust the boundaries make sure to archive the original .TextGrids.

If you have a beep sound with a different frequency, just measure the frequency of the darkest formant. In the following example I have a beep with a dominant frequency at 4602 hz.

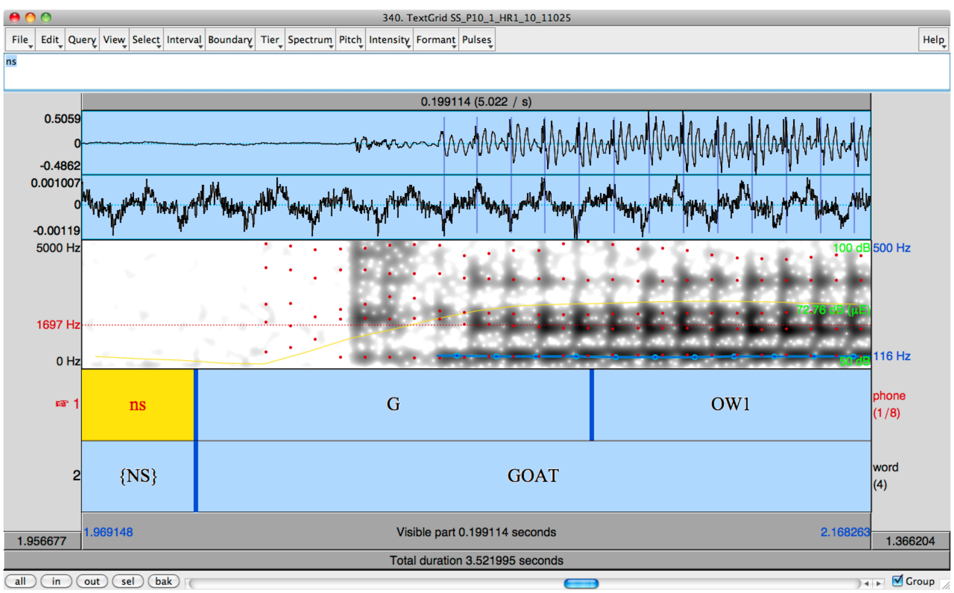


The algorithm has incorrectly identified its onset. I moved the boundary to the onset of the 4602 frequency band.

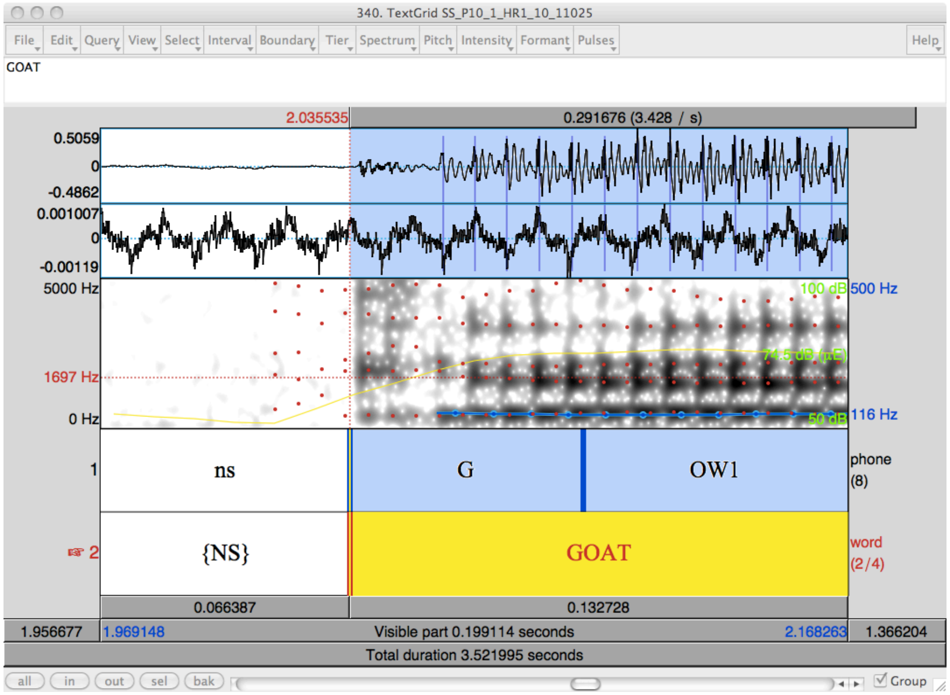


Adjust onset of first word

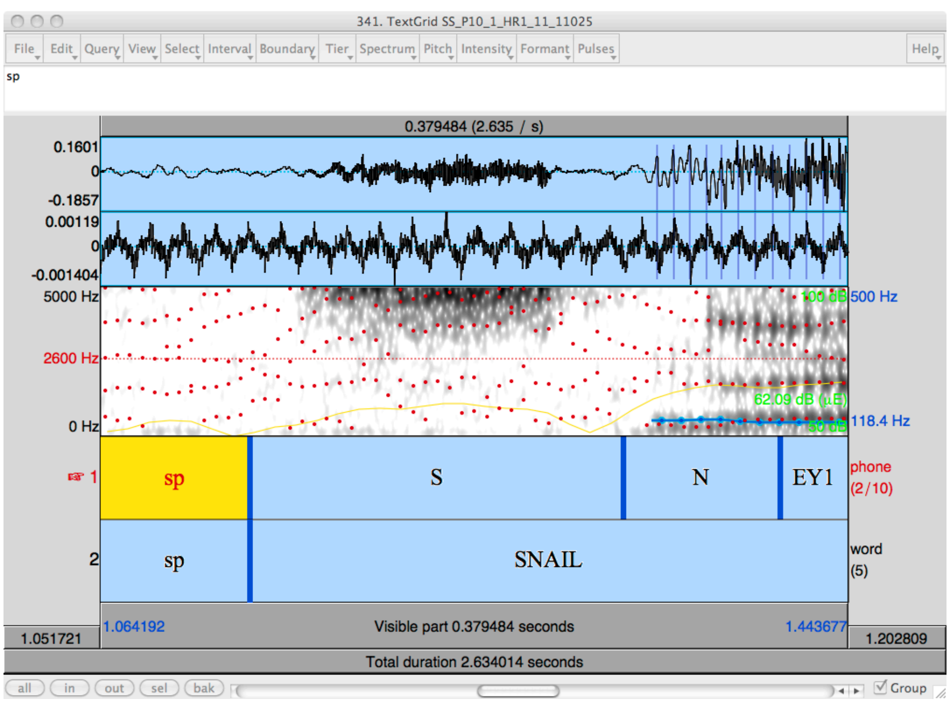
Word boundaries should be identified according to information from both the spectrum and the spectrogram. I use the spectrogram to roughly identify the onset. I then zoom into the region of interest by selecting the area and then hitting the Sel button. For example, in the following screen shot I have zoomed into the onset of the word goat. In the spectrogram you see the appearance of a sudden grey column. This marks the release of the stop consonant g. In the spectrum you see that at the same time point the wave changes form. From almost a flat line, there is a sudden onset of oscillation. For most stop consonants (p,b,t,d,k,g), the appearance of a gray column will be sufficient information to mark the onset of speech. However, you should always consult the spectrum for further confirmatory information for the time of the onset.



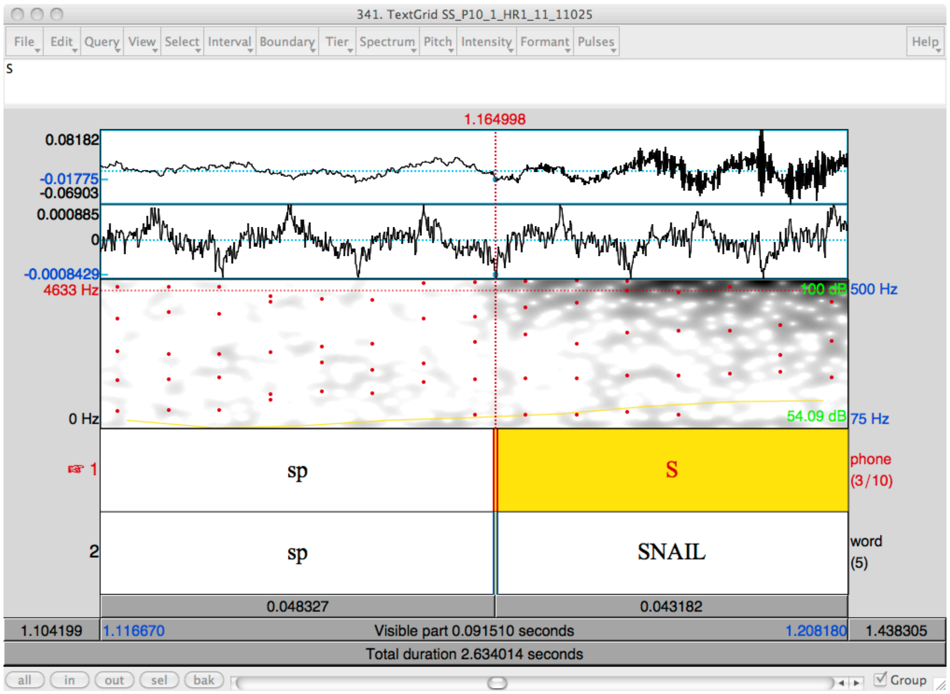
You see that the autosegmentation algorithm has marked the onset of the goat earlier than it should have. We need to correct this by moving the blue mark to the appropriate position I previously identified.



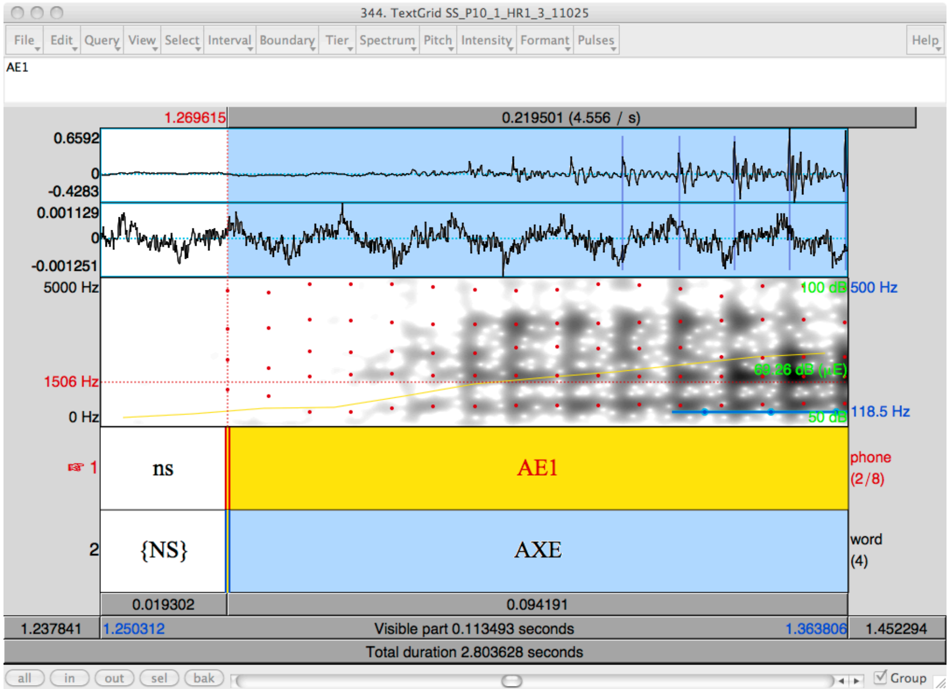
For fricatives such as s,z,f,v,th, and affricates such as ts,tz the onset is marked by the appearance of a band of higher frequency. In the following screen shot I have zoomed into the onset of the word snail. The algorithm has again marked the onset of s earlier than it should have.



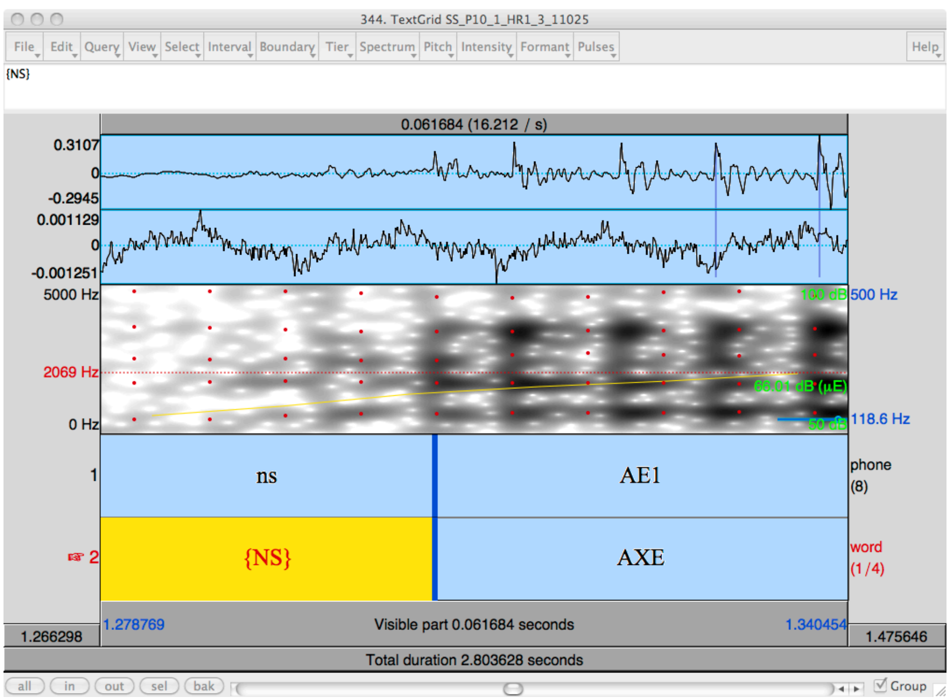
To identify the correct boundary, I zoom into the onset of s even further. I place the boundary where both the spectrogram and the spectrum provide supportive information. The place I placed the onset is where the wave changes form and where the higher frequency band starts to appear.



For vowels the onset is more difficult to identify. It is usually marked by the appearance of the second formant. The spectrum is very useful for the identification of the vowel onset because the onset of the second formant can be somewhat difficult to identify. In the following screenshot I have zoomed into the onset of the word axe.

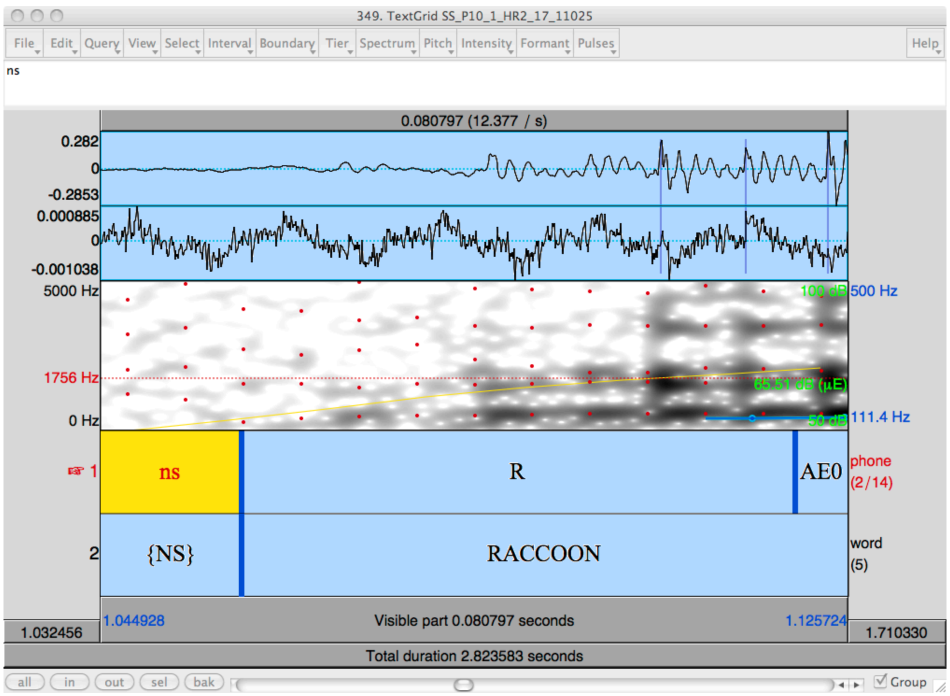


The red dots mark the detected formants by praat. However, they are often incorrect, as in this case. I zoom into the onset even further to have a better look at the spectrum.

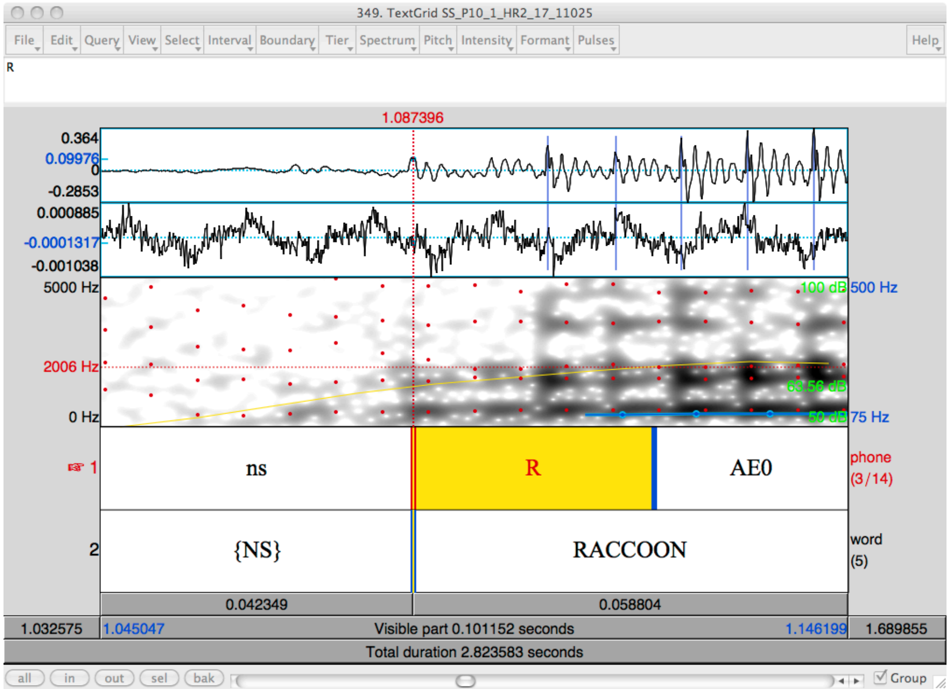


I have marked the onset of *a* at the point where the wave changes form. It is also the place where the wave starts looking more like the oscillations to the right as opposed to the left of it. If you look at the spectrum, the boundary I have placed marks the time point where the second formant is clearly evident.

Liquids and nasals such as r,l,n,m usually look like vowels. Use the same approach as with vowels. In the following screen shot I have zoomed into the onset of raccoon.



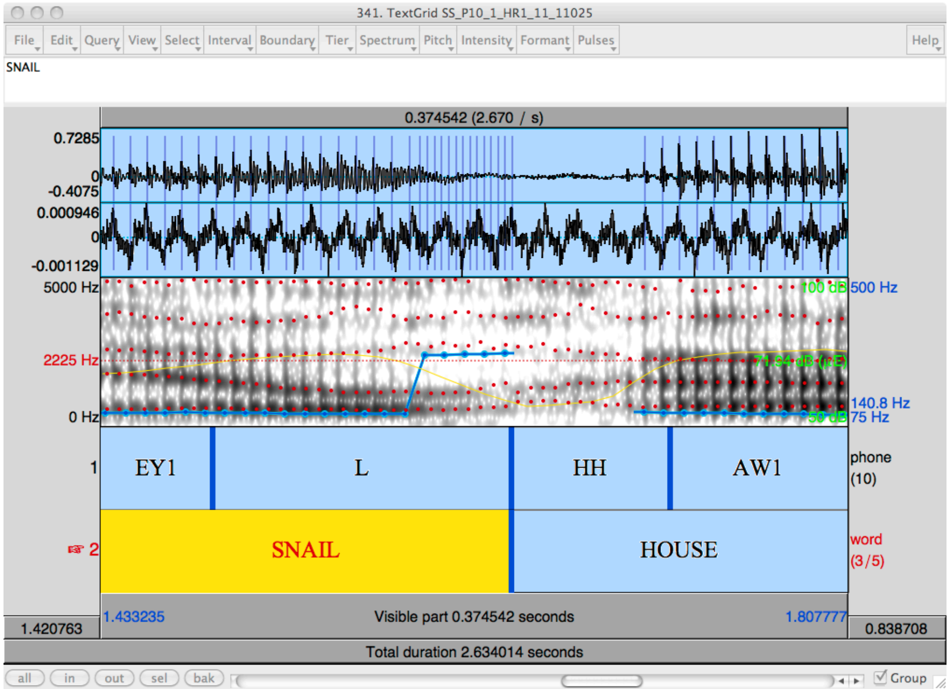
As with all the previous examples, the algorithm has marked an earlier onset of speech. I zoom into the onset a bit further.



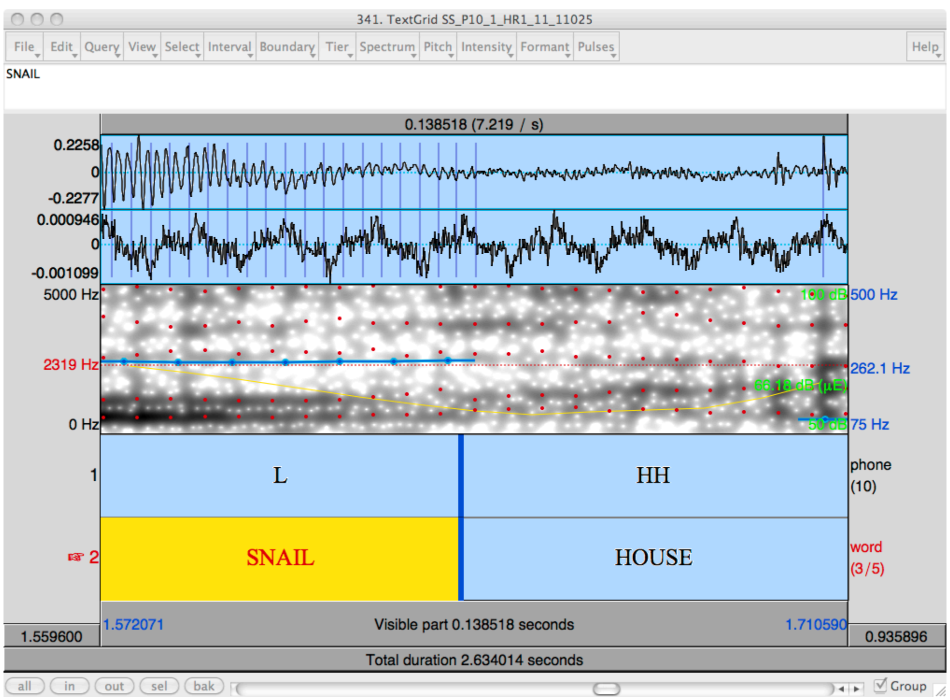
I moved the boundary to where the waveform starts looking more like the oscillations to the right as opposed to the left. The boundary also marks the position of the appearance of a second formant.

Adjust offset of first word

The difficulty with identifying the offset of a word lies in the presence of coarticulation with the onset of an upcoming word. The following picture is an example of coarticulation.

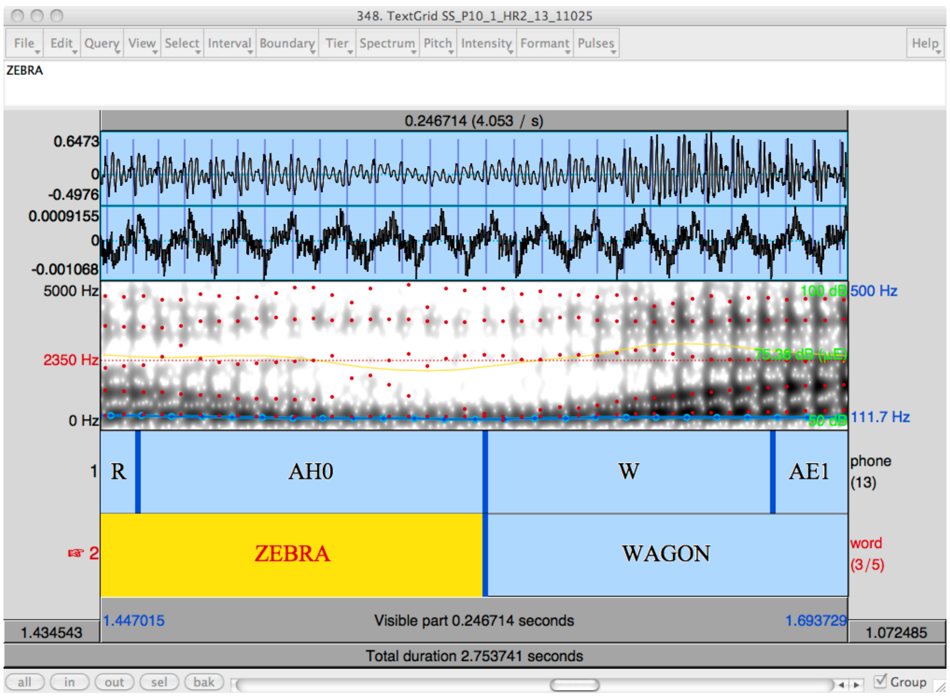


You will notice that there is no clear end of snail because l is coarticulated with the onset of house. To identify the offset of l, I zoom into the region even further.

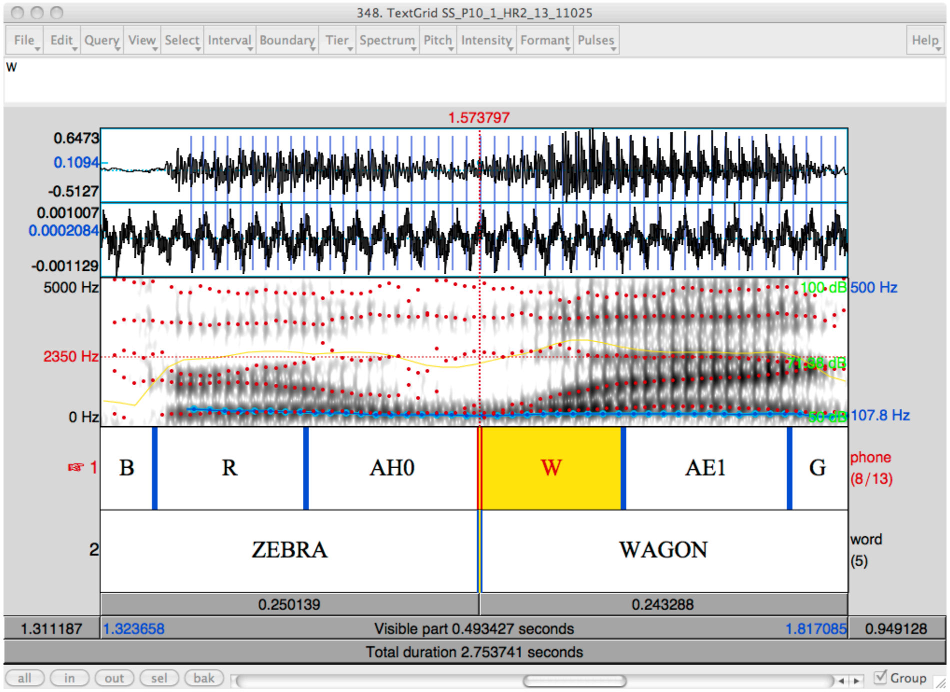


The algorithm seems to have done a pretty good job at identifying where the wave changes form. You will also notice that the boundary coincides with the end of the blue line, which is the representation of the pitch contour. It also coincides with the last vertical blue line in the spectrum which marks praat's identification of the last vocal pulse in the word.

The following screenshot is another example of a difficult case of coarticulation. I have adjusted the boundary according to information in the spectrum and the spectrogram.

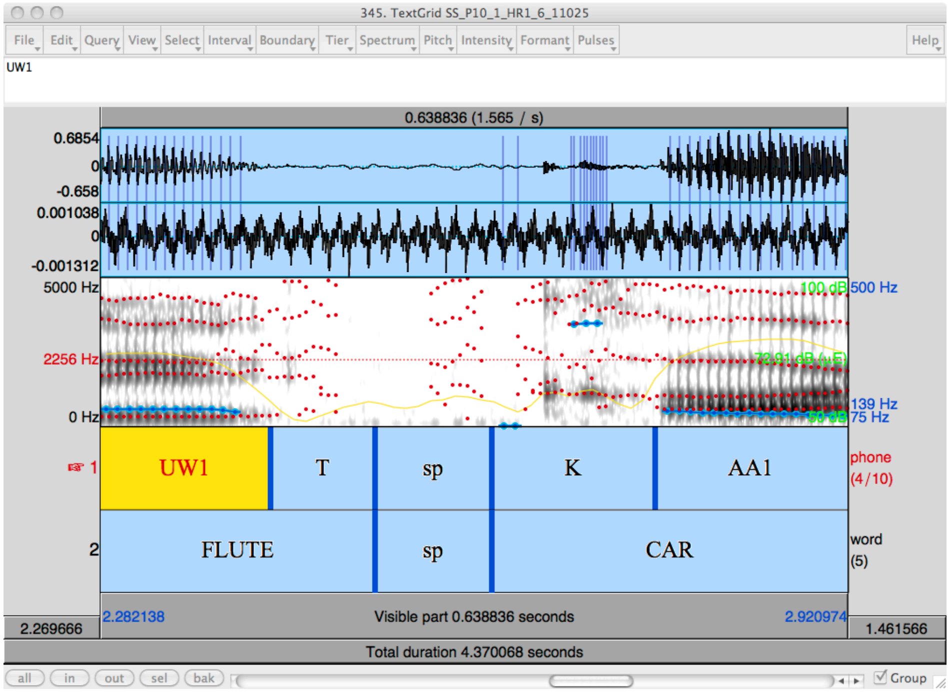


I have placed the boundary where I believe the wave changes form and where the spectrogram indicates the onset of the w. Here is a zoomed out version.

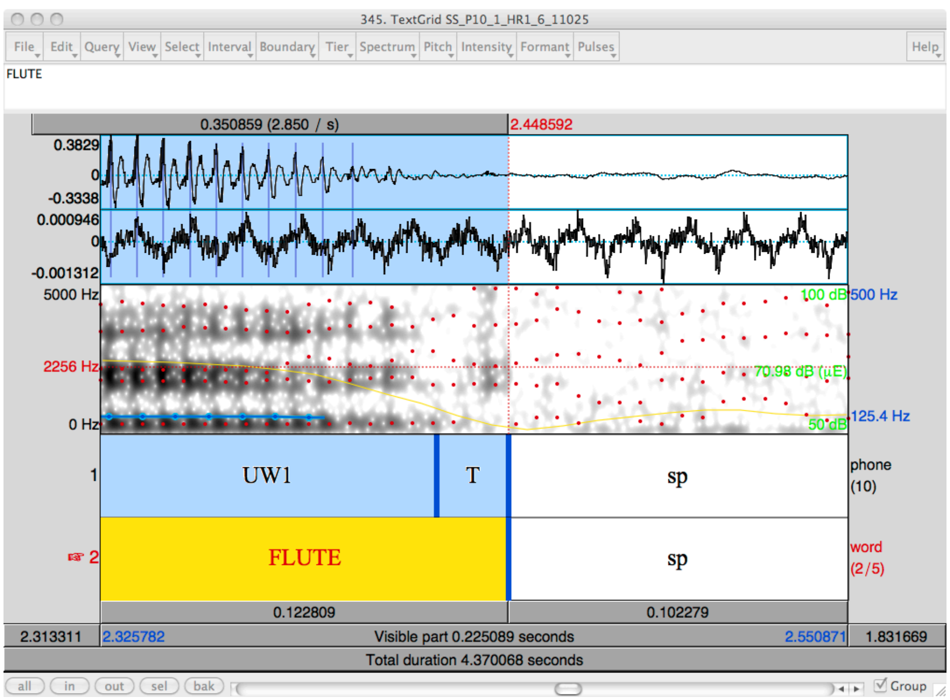


This is a good example of how difficult boundary identification can sometimes be. Chances are that you are going to be a bit off in your detection. As long as you are consistent throughout your boundary adjustment, this will not be a problem.

The identification of the offset of a word can be also difficult because of the presence of a stop at the end of the measured word, or both at the end of the measured word and the upcoming word. The following is an example of the presence of a stop consonant in both the end of the measured word and the beginning of the upcoming word.

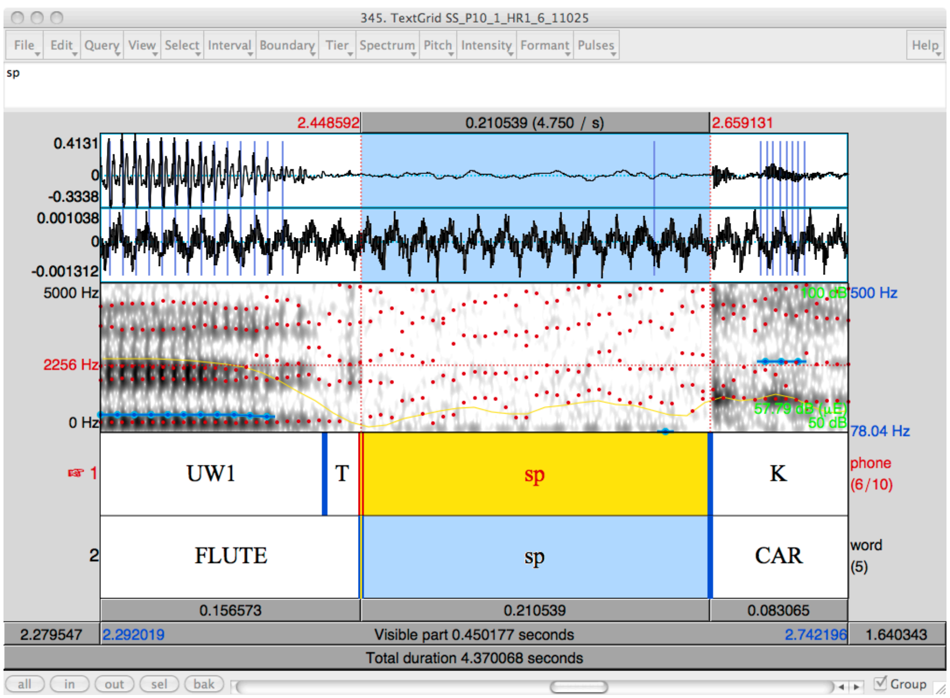


The algorithm has identified a short pause between the two words marked with sp. I zoom into the offset of flute.



I have moved the boundary at the end of the release of the t. There are three pieces of information I used to identify the boundary. In the spectrum, this is the point at which the wave changes form. It is very subtle because stop consonants are not produced by vibrating the vocal chords. Therefore, any information in the spectrum is residual activation from the previous vowel. The second piece of information I use is the presence of a dark band that is probably created by the escape of air from the mouth after the closure of the mouth. The last piece of information is the yellow intensity contour. The contour is at 0 where I have chosen the offset. This signals the end of an acoustic event.

I then adjust the onset of the upcoming word to have an accurate measurement of the duration of the pause.



It is important to note here that the duration of the pause is longer than it objectively is. This is because we can’t identify the onset of the closure of the stop consonant k. The consonant onset can be as long as 80-100ms.

REFERENCES

Alario, F. X., & Costa, A. (2002). Hedging one's bets too much? A reply to Levelt (2002). *Language and Cognitive Processes*, *17*(6), 673-682. doi:10.1080/01690960

Alario, F. X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*(3), 299-319. doi:10.1080/01690960

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments and Computers, 36*(1), 140-55.

Allum, P. H., & Wheeldon, L. (2007). Planning scope in spoken sentence production: the role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791-810. doi:10.1037/0278-7393.33.4.791

Allum, P. H., & Wheeldon, L. (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual-syntactic interface. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1240-1255. doi:10.1037/a0016367

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495-527.

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187-203.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, *56*(4), 521-536. doi:10.1016/j.jml.2006.09.007

Arnold, J. E., & Watson, D. G. (under review). Synthesizing meaning and processing approaches to prosody: Performance matters.

Arnold, J. E., Hudson Kam, C., & Tanenhaus, M. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914-930.

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language*, *76*(1), 28.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*(9), 578-82.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *51*(1), 55-70.

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*(1), 31-56.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H. (2011). LanguageR: Data sets and functions with 'Analyzing linguistic data: a practical introduction to statistics'. *R package version 1.0.*

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. doi:10.1016/j.jml.2007.12.005

Bachoud-Levi, A. C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language*, *39*(3), 331-346.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527-554. doi:10.1080/01690960802299378

Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1289-1302.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*(1), 89-106.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, *28*(1), 14-36.

Bard, E. G., & Aylett, M. P. (2000). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 173-191). Cambridge, MA: MIT Press.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M. P., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language*, *42*, 1-22.

Bates, D., Maechler, M., & Dai, B. (2010). lme4: Linear mixed-effects models using s4 classes [R package version 0.999375-39]. Retrieved from http://CRAN.R-project.org/packages/lme4

Beattie, W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*(3), 201-211.

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*(1), 92-111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., & Gildea, D. (1999). Forms of english function words: Effects of disfluencies, turn position, age and sex, and predictability. In *Proceedings of ICPHS-99* (Vol. 3, pp. 395-398).

Bell, A., Gregory, M. L., Brenier, J. M., Jurafsky, D., Ikeno, A., & Girand, C. (2002). Which predictability measures affect content word durations? In *Proceedings of the ISCA Workshop on Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology* (pp. 1-5).

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, *113*(2), 1001-1024. doi:10.1121/1.1534836

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*(3), 173-194.

Bock, K. J. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1-47.

Bock, K. J. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin and Review*, *3*(4), 395-421.

Bock, K. J., & Cutting, C. J. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*(1), 99-127. doi:10.1016/0749-596X(92)90007-K

Bock, K. J., & Levelt, W. (1994). Language production: Grammatical encoding. *Handbook of psycholinguistics* (pp. 945-984). San Diego, CA, US: Academic Press.

Bock, K. J., Irwin, D. E., Davidson, D. J., & Levelt, W. J. (2003). Minding the clock. *Journal of Memory and Language*, *48*(4), 653 - 685. doi:10.1016/S0749-596X(03)00007-X

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.1.15) [Computer program]. Retrieved August 30, 2009, from http://www.praat.org/

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7), 1044-1098.

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequenas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 274-280. doi:10.1016/j.cognition.2008.07.011

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*(4), 592-609.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *4*(1), 75-87. doi:10.1007/BF01066991

Cholin, J., Dell, G. S., & Levelt, W. J. M. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 109-22. doi:10.1037/a0021322

Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, *50*, 47-61. doi:10.1016/j.jml.2003.08.003

Clark, H. (2002). Speaking in time. *Speech Communication*, *36*(1-2), 5-13.

Clark, H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73-111.

Clark, H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201-242.

Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, *46*(1), 178-198.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 416-431.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*(2), 195-209. doi:10.1016/j.jml.2006.11.001

Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes*, *24*(5), 685-712. doi:10.1080/01690960802351260

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), 892-905. doi:10.1037/a0019446

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283-321.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al (1991). *Psychological Review*, *98*(4), 604-614.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287-314.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123-147.

Eklund, R., & Shriberg, E. (1998). Crosslinguistic disfluency modeling: A comparative analysis of Swedish and American English human human and human machine dialogs. *Proceedings of the International Conference on Spoken Language Processing,* Sydney, November 30 - December 5 (pp. 2631-2634).

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210-233.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, *100*(2), 233-253.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, *22*(8), 1151-1177. doi:10.1080/01690960701461293

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57-84.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296-340.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1187-1199.

Florian J. T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23-62.

Fosler-Lussier, E., & Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in convertional speech. *Speech Communication*, *29*(2-4), 137-158.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, *31*(4), 307-319.

Fowler, C. A, & Housum, J. (1987). Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*, 489-504.

Fowler, C. A., Levy, E. T., & Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language*, *37*, 24-40.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*(15), 1-27.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474-496.

Gahl, S., & Garnsey, M., S. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, *80*(4), 748-775.

Gahl, S., Garnsey, S., Fisher, C., & Matzen, L. (2006). " That Sounds Unlikely ": Syntactic probabilities affect pronunciation. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1334-1339).

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 133-177). London: Academic Press.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. C. T. Walker (Eds), *New approaches to language mechanisms* (pp. 231-255). Amsterdam: North-Holland.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production. Vol.1: Speech and talk* (pp. 177-220). London: Academic Press.

Goldman-Eisler, F. (1968). *Psycholinguistics: experiments in spontaneous speech*. London, New York: Academic Publishing.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649-683.

Gregory, M., Raymond, W.D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (2000). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, *35*, 151-66.

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and semantics 3: Speech acts (pp. 41–58). New York: Academic Press.

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 and Review*, *10*(3), 603-609.

Griffin, Z. M. (2004). The eyes are right when the mouth is wrong. *Psychological Science*, *15*(12), 814-21.

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

Griffin, Z. M., & Oppenheimer, D. M. (2006). Speakers gaze at objects while preparing intentionally inaccurate labels for them. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 943-948.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824-843.

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1256-1274.

Jescheniak, J. D., Schriefers, H., & Hantsch, A. (2003). Utterance format effects phonological priming in the picture-word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 441-454. doi:10.1037/0096-1523.29.2.441

Jurafsky, D., Bell, A., & Girand, C. (2002). The role of the lemma in form variation. In G. Carlos & W. Natasha (Eds.), *Laboratory Phonology 7* (pp. 3-34). Berlin, New York: Mouton de Gruyter.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words : Evidence from reduction in lexical production. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: John Benjamins.

Kahn, J.M., & Arnold, J.E. (in press). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*.

Kawamoto, A. H., Kello, C. T., Higareda, I., & Vu, J. V. Q. (1999). Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset and rime duration. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *25*(2), 362-381.

Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 862-885.

Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, *58*(2), 347-365. doi:10.1016/j.jml.2007.06.002

Kello, C. T. (2004). Control over the time course of cognition in the tempo-naming task. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 942-955.

Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 719-750.

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language*, *48*(1), 207-232.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General*, *129*(3), 340-360.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*(2), 201-258. doi:10.1016/S0364-0213(87)80006-X

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41-104.

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

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Science*, *42*, 1-22.

Levelt, W. J. M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics*, *2*(2), 205–218.

Levelt, W. J. M., & Meyer, A. (2000). Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology*, *12*(4), 433-452.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239-269.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1-38. doi:10.1017/S0140525X99001776

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*(1), 122-142.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172-187.

Longhurst, E. (2006). ExBuilder [Computer program]. Rochester, NY: University of Rochester.

Lupker, S. J. (2003). Mixing costs and mixing benefits in naming words, pictures, and sums. *Journal of Memory and Language*, *49*(4), 556-575. doi:10.1016/S0749-596X(03)00094-9

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 570-590.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. *Psychological Review*, *89*(5), 483-506.

Malpass, D., & Meyer, A. S. (2010). The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects. *Journal of experimental psychology: Learning, Memory, and Cognition*, *36*(2), 523-37. doi:10.1037/a0018522

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69-89.

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., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(6), 1146-1160.

Meyer, A. S., & van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin and Review*, *7*(2), 314-9.

Meyer, A. S., Ouellet, M., & Haecker, C. (2008). Parallel processing of objects in a naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 982-987. doi:10.1037/0278-7393.34.4.982

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., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition*, *66*(2), B25 - B33.

Meyer, A. S., Belke, E., Hacker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language*, *57*(2), 210-231. doi:10.1016/j.jml.2006.10.005

Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *The Journal of the Acoustical Society of America*, *95*(3), 1603-1616. doi:10.1121/1.408547

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273-281. doi:10.1080/17470216508416445

Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2010). Phonological advance planning in sentence production. *Journal of Memory and Language*, *63*(4), 526-540. doi:10.1016/j.jml.2010.07.004

Owren, M. J. (2008). GSU Praat Tools: scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods*, *40*(3), 822-829.

Quene, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: a tutorial. *Speech Communication*, *43*(1-2), 103-121.

Roelofs, A. (1998). Rightward incrementality in encoding simple phrasal forms in speech production: Verb-particle combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 904-921.

Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. *Journal of Memory and Language*, *57*(2), 232-251.

Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1580-1598.

Roelofs, A. (2011). Attention demands of spoken word planning: a review. *Frontiers in Psychology*, *2*(November), 1-14. doi:10.3389/fpsyg.2011.00307

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217-236.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research*, *35*(2), 189-213. doi:10.1007/s10936-005-9011-6

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123-164). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Schriefers, H. (1999). Phonological facilitation in the production of two-word utterances. *European Journal of Cognitive Psychology*, *11*(1), 17-50.

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*(1), 86-102.

Shields, L. W., & Balota, D. A. (1991). Repetition and associative context effects in speech production. *Language and Speech*, *34*(1), 47-55.

Shriberg, E. (1999). Phonetic consequences of speech disfluency. *Proceedings of the International Conference on Phonetic sciences,* San Francisco (pp. 619-622).

Shriberg, E. (2001). To "errr" is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*(01), 153-169. doi:10.1017/S0025100301001128

Slavin, S. (Lancaster U.) (2006). PsyScript. [Computer program]. Retrieved September 2009, from http://www.psych.lancs.ac.uk/software/psyScript.html

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174-215.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments and Computers*, *28*(4), 516-536.

Sternberg, S., Wright, C., & Knoll, R. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 507–534). Englewood Cliffs, NJ: Erlbaum.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.) *Information processing in motor control and learning* (pp. 117-152). New York: Academic Press.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing ‘‘given’’ information: Effects of persistence of grammatical function and surface position. *Language and Speech, 37,* 125-145.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, *1*(2), 147-165. doi:10.1515/LANGCOG.2009.008

Van der Meulen, F. F., Meyer, A., & Levelt, W. J. M. (2001). Eye movements during the production of nouns and pronouns. *Memory and Cognition*, *29*(3), 512-521.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423-40. doi:10.1037/a0018619

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*(3), 1548-1557. doi:10.1016/j.cognition.2007.06.009

Wheeldon, L. R., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*(3), 356-381. doi:10.1006/jmla.1997.2517

Wheeldon, L. R, & Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical word. *Cognition*, *85*(2), B31-B41.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*(3), 311–334.

Wheeldon, L. R., & Monsell, S. (1992). The Locus of repetition priming of spoken word production. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *44*(4), 723-761.

Wheeldon, L. R., Smith, M. C., & Apperly, I. A. (2011). Repeating words in sentences: Effects of sentence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1051-64. doi:10.1037/a0023435

Yuan Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics ‘08*, Paris (pp. 5687-5690).