Acoustic Representations of Segmental and Metrical Encoding in Speech Production

by

Brett R. Myers

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

INTERDISCIPLINARY STUDIES:
PROSODY, LANGUAGE AND MUSIC

May 10, 2019

Nashville, Tennessee

Approved:

Reyna L. Gordon, Ph.D.

Duane G. Watson, Ph.D.

Cyrille L. Magne, Ph.D.

Miriam D. Lense, Ph.D.

Stephen M. Camarata, Ph.D.

To metrical variations:
may they lead us through moments of stress
and always keep us on our feet.

ACKNOWLEDGMENTS

How do you sum up all of the gratitude that you have for all of the incredible people who contributed to a successful dissertation? Well, I suppose two pages at the beginning will do, but really these people deserve much more than that.

First and foremost, I am extremely indebted to my advisors Duane and Reyna, or as I collectively call them—Duayna. Thank you both for your encouragement, patience, and guidance through these years. Reyna, your enthusiasm for science is inspiring, and I am grateful for the time, energy, and careful attention to detail that you've given to our work together. You were brave enough to navigate the land of interdisciplinary studies with me, and without that, I would not be here today. Duane, I have learned so much from you—thank you for being a prime example of an impressive, intelligent researcher. I owe it to you for getting me interested in this topic; it's amazing how a simple conversation in your office can grow into a series of experiments. I really lucked out to have you both supporting me, and I couldn't have asked for a better pair of mentors.

I am also very grateful to my committee members—Stephen, Miriam, and Cyrille—for the many thoughtful conversations and for pushing me to always think critically about my research. I also owe a special thanks to Richard Hoover, John Rieser, and Dan Ashmead for steering me in the right direction through the grad school maze and helping me get to where I ultimately needed to be.

The Department of Otolaryngology and my family at the Vanderbilt Voice Center deserve a major salute for believing in me enough to let me venture out on this journey while still working in the clinic. Thanks especially to Tom, Gaelyn, Jenny, Cristen, Jenn,

Morgan, Rachel, Paige, Emily, and Amy for cheering me on and for gently nudging me with the occasional “Have you finished yet?”

As a student with two advisors, I had the good fortune of being involved with two lab groups, and I am so thankful for all of the people I’ve had the pleasure to work with in those labs. From the Communication and Language Lab (and its affiliates), I appreciated the thoughtful theoretical conversations that sparked many ideas in this research with Cassie, Andres, Esteban, Sarah, Caoimhe, and Nick. Also thanks to the research assistants who suffered many hours of data collection and analysis: Michael, Joe, Yujia, Yiran, Jordan, Marcus, and Zisu. From the Music Cognition Lab, I will always cherish the camaraderie, joint music making, and happy hours especially with Kate, Chloe, Alex, Karli, and Rachana. Most of all, thanks to Allison, who always makes me laugh out loud actually—everyone needs a friend like her when dissertating.

Finally, thanks to my family and friends for some much needed work-life balance. To my mom, dad, sister, and the entire extended crew: thank you for the unending support and for always believing in me. To my friends who kept me sane during all the craziness, especially Justin, Dora, Jacob, Abigail, Nathaniel, and everyone in the theatre squad: thank you for the encouragement, the laughter, and for giving me a creative outlet when my brain would otherwise implode.

I’d also like to acknowledge Friedlieb Ferdinand Runge, the chemist who discovered caffeine in 1819—as well as my local baristas who always get my coffee just right—their contributions to this science cannot be overstated.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

LIST OF FIGURES vii

Chapter

1. Introduction 1

   Acoustic Features of Prosody 1
   Acoustic Prominence: Word Duration 3
   Acoustic Prominence: Word Stress 7

2. Experiment 1 11

   Methods 11
   Results 16
   Discussion 20

3. Experiment 2 22

   Methods 22
   Results 27
   Discussion 33

4. Experiment 3 38

   Methods 38
   Results 41
   Discussion 44

5. General Discussion 46

   Implications and Future Directions 46
   Conclusions 55

REFERENCES 56

Appendix

A. Stimulus List for Experiment 1 70
B. Praat Coding Manual 71
C. Stimulus List for Experiment 2 73
D. Speech Errors Coding Manual 74
E. Stimulus List for Experiment 3 75
F. Summary of Statistical Findings 76

LIST OF TABLES

Table Page

1. Fixed Effects Estimates for Experiment 1 18

2. Fixed Effects Estimates for Experiment 2 30

3. Speech Errors in Experiment 2 32

4. Fixed Effects Estimates for Experiment 3 43

LIST OF FIGURES

Figure Page

1. Models of Metrical Planning 8
2. Example of Event Description Task 13
3. Example of Praat Segmentation in Experiment 1 15
4. Results for Experiment 1 18
5. Example of Praat Segmentation in Experiment 2 26
6. Results for Experiment 2 29
7. Distribution of Speech Errors in Experiment 2 31
8. Results for Experiment 3 42

CHAPTER 1: INTRODUCTION

Phonological encoding—the process of selecting and organizing phonemes for word production—is thought to consist of two main branches: segmental and metrical spell-out (e.g., Levelt et al., 1999). It has been proposed that segmental spell-out occurs in a serial sequence (Sevald & Dell, 1994) that requires processing time to complete (Watson et al., 2015); however, the nature of metrical spell-out is less obvious. This paper investigates the role of metrical stress in the phonological encoding process through the course of three empirical studies. The first chapter presents an overview of the role of prosody in spoken language, as well as word production studies that have provided a basis for theories of phonological encoding. The following chapters outline hypotheses, methods, and findings of three experimental designs. Based on the results of these experiments, we propose a novel theory of segmental and metrical encoding.

1.1 ACOUSTIC FEATURES OF PROSODY

Prosody is a suprasegmental linguistic function that provides stress, intonation, and rhythm to an utterance, and it operates at multiple levels from phonemes and syllables to words and phrases (Kunert & Jongman, 2017; Dahan et al., 2002; Pitt & Samuel, 1990). Prosodic cues of an utterance indicate a speaker's affect and intent (Scherer, 1986) as well as emphasis, sarcasm, and more nuanced emotional states (Zentner et al., 2008; Coutinho & Dibben, 2013). Take for instance a quote from George Bernard Shaw: "The trouble with her is that she lacks the power of conversation but not the power of speech." We may interpret the speaker's attitude toward the subject by the way he may emphasize *conversation* and *speech*, as well as tonal characteristics of his voice. Certain prosodic cues can be interpreted cross-culturally even in an unfamiliar language due to the acoustic gestures in an utterance (Scherer et al., 2001; Thompson & Balkwill, 2006). For example, a non-Spanish-speaking listener may hear, “Cómo estás?” and recognize that this is a question–simply based on pitch inflection. Prosodic features interact with semantics, syntax, and pragmatics to express the meaning of a spoken message. Prosodic perception in infancy leads to vocabulary and grammar development (Gervain & Werker, 2013; Nazzi & Ramus, 2003; Soderstrom et al., 2003), and it is believed to impact later linguistic abilities, literacy, and social interactions (e.g., Gordon et al., 2015; Holt et al., 2017; Goswami et al., 2010; Grossman et al., 2010). Additionally, the prosodic contour of speech may be a fundamental feature that drives neural entrainment to speech (Myers et al., 2019).

The prosodic fluctuations in an utterance are conveyed through acoustic correlates such as duration, amplitude, and fundamental frequency. As any of these parameters changes, it influences the expression of stress, intonation, and rhythm of the spoken message (Fletcher, 2010; Lehiste, 1970). Hermione Granger emphasizes the importance of prosodic expression in *Harry Potter and Sorcerer’s Stone*, as she cleverly explains to Ron that the magical spell *wingardium leviosa* is pronounced “levi-OH-sa” not “levio-SAH.” Here she uses syllable lengthening, pause boundaries, pitch inflection, and vocal intensity to get her point across. All of these features contribute to the acoustic prominence of syllables and words (Greenberg et al., 2003), which facilitates segmenting speech into syllable components and disentangling word boundaries (Jusczyk et al., 1992; Leong & Goswami, 2015). In English, speakers may adapt the acoustic structure of a word in any number of ways, and the next section outlines some hypotheses for why these adaptations may occur.

1.2 ACOUSTIC PROMINENCE: WORD DURATION

One facet of acoustic prominence that changes dynamically throughout production is word duration. Word duration can be tailored—through reduction or lengthening—to fit certain parameters of the discourse. For instance, speakers tend to reduce the duration of words that were recently uttered (e.g., Fowler & Housum, 1987; Breen et al., 2010; Buxó-Lugo et al., 2018). When the titular character in *Richard III* yells out, "A horse, a horse! My kingdom for a horse!," the final *horse* will likely be shorter than the first two. On the other hand, speakers tend to lengthen a word that shares partial segmental overlap with previously uttered words (e.g., Sevald & Dell, 1994; O'Seaghdha & Marin, 2000; Yiu & Watson, 2015); when Peter Quince in *A Midsummer Night's Dream* delivers his line, "With bloody blameful blade, he bravely broached his boiling bloody breast," *blade* will likely be longer than usual because of the segmental overlap with the preceding *blameful* (not to mention the overall production difficulty due to alliteration throughout). Reduction and lengthening have been studied extensively, and there are generally two schools of thought to explain this phenomenon.

1.2.1 Audience-Design Hypothesis

A common hypothesis is a communicative account that suggests we adapt duration for the benefit of the listener (e.g. Jaeger, 2010; Aylett & Turk, 2004; Bard et al., 2000). For clarity in discourse, the speaker will hyperarticulate and thereby lengthen a potentially confusable word to avoid contextual ambiguity (Buz et al., 2016). Buz et al. (2016) found that speakers hyperarticulated when a simulated partner experienced confusion; speakers used interlocutor feedback to evaluate and clarify their message. Furthermore, they found that speakers adapted subsequent productions based on unsuccessful trials—that is, in anticipation of confusion, they hyperarticulated new words to facilitate successful communication. Conversely, if a word has already been mentioned, the speaker can afford to reduce the care of articulation for that word since there is greater likelihood that the word will be recognized (e.g., Aylett & Turk, 2004).

The Smooth Signal Redundancy Hypothesis (Aylett & Turk, 2004) suggests that redundancy should be evenly distributed throughout an utterance, and speakers use prosodic prominence to maintain an inverse relationship between linguistic and acoustic redundancy. In other words, speakers will lengthen words that are new and shorten words that are repeated for the sake of smooth signal redundancy. This suggests that speakers adapt their speech to facilitate information transfer for their communicative partners (Buz et al., 2016; Cohen Priva, 2015; Seyfarth, 2015; Pate & Goldwater, 2014; Tily & Kuperman, 2012; Galati & Brennan, 2010; Aylett & Turk, 2004).

1.2.2 Internal Production Hypothesis

Another explanation is that durational adjustments occur as a result of production-internal mechanisms (e.g., Dell, 1986; Baese-Berk & Goldrick, 2009); that is, word duration changes for the benefit of the speaker rather than for an audience. This theory suggests that reduction or lengthening reflect the amount of time needed for the production system to produce a linguistically easy or complex word (Bell et al., 2009;

Kahn & Arnold, 2012). While the audience-design account claims that repeated words are reduced to match the information load, the production approach claims that repeated words are reduced because they have already been primed in the production system (Pickering & Garrod, 2004). Both of these systems likely contribute to durational effects (Arnold & Watson, 2015), but the remainder of this study will focus on production processing mechanisms.

1.2.2.1 Serial Ordering of Phonemes

Phonological encoding is the process of retrieving the form of a word by selecting and ordering phonemes to be used during production. Some models suggest that the phonological encoding process occurs incrementally while the word is being articulated, which affects word duration (e.g., Sevald & Dell, 1994; O'Seaghdha & Marin, 2000). These models suggest that the ordering of speech sounds is a serial sequencing process that takes time—harder words take more time to complete, easier words take less time—and this has a direct impact on the length of a word (Watson et al., 2015).

In a seminal study, Sevald and Dell (1994) asked participants to repeat two-word phrases as quickly as possible, and they found that word-initial overlap (e.g., *pick-pin*) yielded slower production than word-final overlap (e.g., *pick-tick*). According to their model, phonemes are accessed in the order in which they appear in a word, and activation of one phoneme triggers activation for the next phoneme. Therefore, when word onsets overlap (*pi-*), both words in the pair become activated (*pick* and *pin*). Ambiguity arises because both words were recently activated, and the production system must decide which phoneme (*-k* or *–n*) should follow the initial segment. However, this competition does not arise in overlapping offsets (*-ik*) because the appropriate phoneme sequence is established from the onset (*p-* or *t-*). That is, the words (*pick* or *tick*) are differentiated starting at the first phoneme. This model explains the consequences of word lengthening and reduction; namely, similar phoneme sequences create interference at the phonological level (Watson et al., 2015), and conversely repeating a given word can facilitate a speedier activation of that phonemic sequence (Kahn & Arnold, 2012; Lam & Watson, 2010).

Lengthening and reduction have also been measured in word priming picture description tasks. When repeating the name of an image, phoneme retrieval is easy for the production system and will result in word reduction (Kahn & Arnold, 2012; Lam & Watson, 2010). The same reduction has been observed with homophone word pairs, where participants name two images with the same phonemic sequence, and duration of the second word is reduced (Jacobs et al., 2015). This indicates that the effect is likely occurring at an acoustic-phonetic level rather than a lexical-semantic level of speech processing. Repetition reduction was also observed to a higher degree in overtly spoken words compared to words in inner speech (Kahn & Arnold, 2015; Jacobs et al., 2015). Furthermore, reduction or lengthening can occur when either the speaker or someone else says the prime word (Kahn & Arnold, 2015; Buxó-Lugo et al., 2018), suggesting that the acoustic form of a word is enough to prime phonological encoding. Together, these findings demonstrate that lengthening and reduction are the result of priming the production system with auditory representations of word forms, rather than other types of representations (i.e., lexical, phonological, articulatory).

1.3 ACOUSTIC PROMINENCE: WORD STRESS

An important yet less investigated component of acoustic prominence is the metrical structure of a word. Meter refers to the pattern of stress given to consecutive syllables. Syllables may be stressed by accentuating pitch, intensity, and durational variations (e.g., Fry, 1955; Lieberman, 1960), which is thought to facilitate parsing speech into meaningful linguistic segments for the listener (Kunert & Jongman, 2017; Pitt & Samuel, 1990). Meter can also play a role in distinguishing between two phonologically similar words, such as *DES-ert* (a barren landscape) versus *dess-ERT* (a tasty treat). Disyllabic words such as these can be categorized as trochees—with stress on the first syllable—or iambs—with stress on the second syllable. While native speakers of English establish a metrical structure without much difficulty, it is altogether unclear how we build this structure during phonological encoding.

Current linguistic theory suggests that segmental and metrical information exist on separate representational levels and are retrieved independently in word-form generation (e.g., Goldsmith, 1990; Kenstowicz, 1994). Levelt and colleagues (1999) suggest that a metrical frame—consisting of the number of syllables and the location of stress—is retrieved during phonological encoding, and then segmental units are associated to the frame in syllabification. Roelofs & Meyer (1998) tested this theory as part of their Word-form Encoding by Activation and Verification (WEAVER) model (Roelofs, 1997). They used a series of implicit priming paradigms in which participants learned sets of word pairs; then the participants would see a prime word and were instructed to say the matching target word as quickly as possible. Their results showed a priming effect in which words with both segmental and metrical overlap were produced in a shorter reaction time, suggesting that speakers could access the target words more quickly when they shared these features. They did not find an effect with only segmental or metrical overlap alone, which led to their claim that segmental and metrical spell-out occur in parallel and at the same time. However, one could argue that these features are closely connected, which explains the larger effect when they co-occur. Figure 1 presents an example schema of this theory.

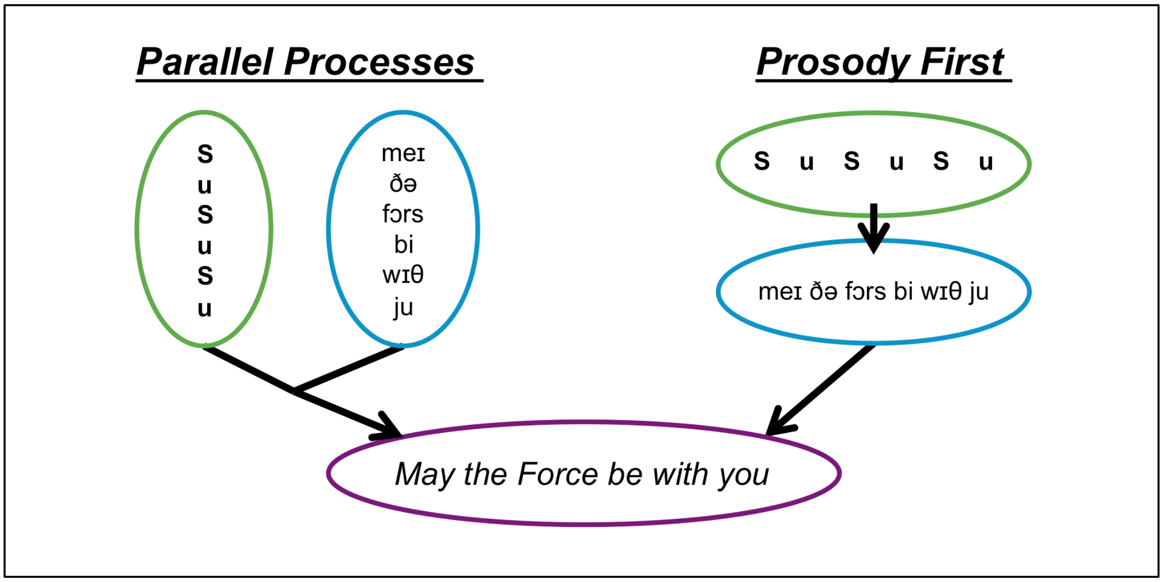


Figure 1. Models of metrical planning. In the parallel processes model, metrical and segmental information are planned simultaneously. In the prosody-first model, metrical information is planned first, and segments are then ordered within the metrical framework. Here, **S** denotes a stressed syllable and **u** denotes an unstressed syllable.

Other models have proposed a prosody-first approach to phonological encoding (e.g., Keating & Shattuck-Hufnagel, 2002), in which the prosodic structure of an entire utterance is planned prior to retrieving word segments. In this model, locations of syllable stress are determined first, and then segments are assigned to the appropriate positions within that metrical frame. Evidence for the prosody-first model comes from analyses of speech errors. Speakers often make errors by misplacing sound segments or recruiting the wrong segment for a word, such as “well-boiled icicle” (*well-oiled bicycle*) or “Is the bean dizzy?” (*Is the dean busy?*). Misplaced segments generally maintain their position within a syllable (e.g., Boomer & Laver, 1968); that is, an onset exchanges for an onset, nucleus for nucleus, and coda for coda (MacKay, 1970; Motley, 1973; Nooteboom, 1969; Shattuck-Hufnagel, 1983, 1987; Stemberger. 1982), suggesting that there is a pre-determined metrical outline that is independent of segments. In addition, when speakers produce segmental errors, the overall stress pattern of the utterance is typically preserved (Berg, 1990; Shattuck-Hufnagel & Turk, 1996; Fromkin, 1971). Lastly, there is work showing that stress in the initial syllable (i.e., trochaic stress pattern) creates a stable architecture for segmental spell-out, as evidenced by fewer segmental errors in trochees versus iambs (Beirne & Croot, 2018; Aichert et al., 2016; De Jong, 1995; Sulpizio et al., 2015).

There is ample evidence that segmental encoding occurs incrementally (e.g., Meyer, 1990, 1992; Shattuck-Hufnagel, 1992; Sevald & Dell, 1994; Wheeldon & Levelt, 1995; Yiu & Watson, 2015), and some have posited that metrical encoding is spelled out in a similar fashion (Schiller et al., 2006; Levelt et al., 1999). If meter is established in serial order like phonemic segments, one may expect to see word lengthening when words share overlapping metrical frames because competition for the initial stress value would slow production. To further investigate the organization of segmental and metrical spell-out, we conducted a series of three experiments. Experiments 1 and 3 were designed as picture description priming tasks (as in Yiu & Watson, 2015), and Experiment 2 was a minimal pair repetition task (as in Sevald & Dell, 1994). Experiment 1 investigated word lengthening when prime-target pairs shared segmental overlap with and without metrical overlap. If metrical planning engages the same types of encoding mechanisms as segmental planning (Roelofs & Meyer, 1998), we should see similar overlap-driven lengthening effects, with longer productions of a target word when the prime shares the same metrical structure. Experiment 2 tested metrical structure as a serial ordering process using repeated word pairs with and without metrical overlap. If metrical planning is a serial process like segmental planning (Sevald & Dell, 1994), we should see similar lengthening effects with metrical overlap. In Experiment 3, we manipulated both segmental overlap and metrical overlap independently to directly test whether segmental and metrical planning engage independent processes (Roelofs & Meyer, 1998) or whether they share the same underlying representation. Together these experiments systematically examine whether metrical and segmental spell-out occur independently and in a similar manner during phonological encoding.

CHAPTER 2: EXPERIMENT 1

The first experiment was designed to examine the role of meter in a word lengthening paradigm. In this experiment, we used a picture description task to elicit productions of prime-target word pairs that shared segmental overlap with and without metrical overlap. Previous evidence has shown that phoneme selection occurs serially (e.g., Sevald & Dell, 1994; Yiu & Watson, 2015), and words lengthen when they share initial phonological segments. If meter is planned in a similar manner to segmental spell-out (as in Roelofs & Meyer, 1998), then we would expect the word lengthening effect to increase when words share both segmental and metrical overlap. That is, we predicted that word pairs with segmental and metrical overlap would lengthen more than word pairs with segmental overlap alone due to the increased complexity at the phonological encoding level.

2.1 METHODS

2.1.1 Participants

Sixty-nine healthy adults (age range: 18-27, *M* = 20.3 years, *SD* = 2.4, 51 female) participated in this study. Participants were native speakers of English recruited from the Vanderbilt University Psychology Department subject pool, and they either received course credit or $10 for participating in the study. All participants provided written informed consent in accordance with the Vanderbilt University Institutional Review Board.

2.1.2 Materials

A set of 144 color images was selected from the Snodgrass and Vanderwart (1980) dataset (Rossion & Pourtois, 2001) and Clipart. A subset of 72 images served as the critical items, and the remaining 72 images were filler items. Critical trials consisted of 18 targets and 54 primes. Prime-target pairs were arranged into three conditions:

1. **Segmental & Metrical Overlap:** The *candy* shrinks. The *candle* flashes.
2. **Segmental Overlap Alone:** The *canteen* shrinks. The *candle* flashes.
3. **Control:** The *giraffe* shrinks. The *candle* flashes.

In the two experimental conditions, the prime-target pairs had segmental overlap for their initial segments, and the meter of the words either matched (1) or did not match (2). In the control condition (3), the prime-target pairs had no segmental overlap and had non-matching meter.

A Latin square design yielded three counterbalanced lists of items, such that each participant was presented with 18 critical prime-target pairs. Each list had six critical pairs for each of the three conditions. An equal number of trochees and iambs were used as critical targets. In addition, participants were exposed to 38 non-critical pairs, drawn from the filler items, for a total of 56 trials in the experiment. Trials were randomized for each participant. See Appendix A for the list of critical prime-target pairs.

2.1.3 Audio Recording

Participant responses were recorded via a head-mounted microphone at a sampling rate of 44,100 Hz. Participants were instructed to speak directly into the microphone as they described the events on the computer screen.

2.1.4 Procedure

Participants completed the experiment on a Mac computer in Matlab using the CogToolbox (Fraundorf et al., 2014) and Psychophysics Toolbox 3 (Kleiner et al., 2007). Participants first completed a training task to learn the names of potentially difficult to name items (e.g., Trude & Brown-Schmidt, 2012). During training, items were displayed in the center of the screen with the intended label at the top of the screen, and participants recited the label aloud. They were encouraged to use these names during testing.

Following item training, participants received instructions for the experiment. For each trial, four images were displayed equidistant around the center of the screen (see Figure 1). One image—the prime—would shrink, and participants described the action. Then another image—the target—would flash, and participants described the action. Events occurred in the same order for all trials (i.e., shrinking then flashing). Trials were randomized and separated into three blocks, allowing participants to take a break between blocks as needed.

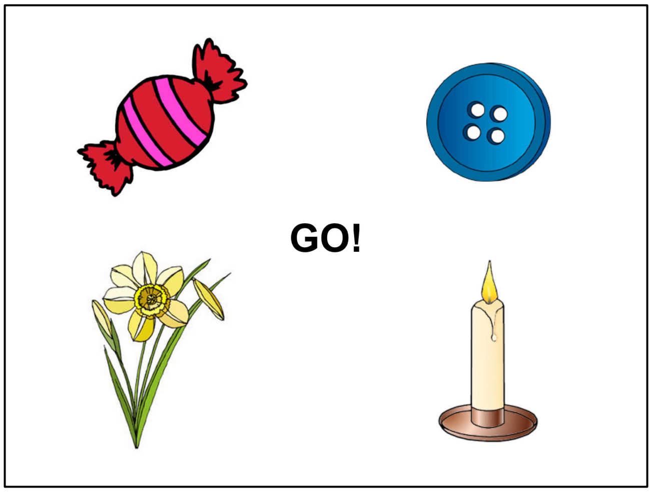


Figure 2. Example of event description task. The images for *candy* and *candle* form a critical prime-target pair, and *flower* and *button* are filler items.

2.1.5 Acoustic Analysis

Speech recordings were analyzed in Praat (Boersma & Weenink, 2017), using manual segmentation to code the start and end times of target words. Each trial was segmented in isolation using spectrographic and waveform information, and coders were blinded to the experimental condition of trials. Target words were segmented such that they were not identifiable as anything other than the targets.

Coders were trained on segmenting the target words within the recorded utterance. To do this, they used a Praat script that opened a graphical user interface (GUI) in which they specified which trial to segment. A spectrographic and waveform display of the audio recording appeared, along with an annotation tier. The spectrogram display utilized the Praat features of Formant and Pitch, which were helpful in determining word boundaries. Coders then listened to the trial to identify the general location of the target word, and they zoomed in to the region of the word to get a clear view of the spectrogram and waveform. They used both audio and visual cues to locate the target word. They were trained on identifying phoneme-specific spectrographic information (e.g., the burst of energy in a plosive, the formant characteristics of vowels), and particular attention was given to distinguishing between the final consonant of a target word and the initial /f/ of "flashes" (which always followed the target word). Coders marked the start point of a target word by placing a boundary where the initial consonant energy began; this did not include any preceding silence or any part of the preceding word. Coders marked the end point of a target word by placing a boundary where the final consonant energy stopped; this did not include subsequent silence or any part of the following word. They played back the target word in isolation to determine if they accurately identified the target so that the full length of the word was captured and nothing other than the word existed within the boundaries. After listening, they adjusted the boundaries as needed to ensure that they made an accurate segmentation. See Figure 2 for an example. When they were finished with a trial, they clicked “Continue” in the GUI, which automatically saved their annotation as a .TextGrid file and opened the next trial. See Appendix B for the Praat coding manual used in data analysis.

The duration for each target word was extracted from all trials using a Praat script that captures the amount of time between the two boundaries on either side of the target word. This script provided output signifying participant, condition, target word, and word duration for each trial, and this dataset was used in statistical analysis.

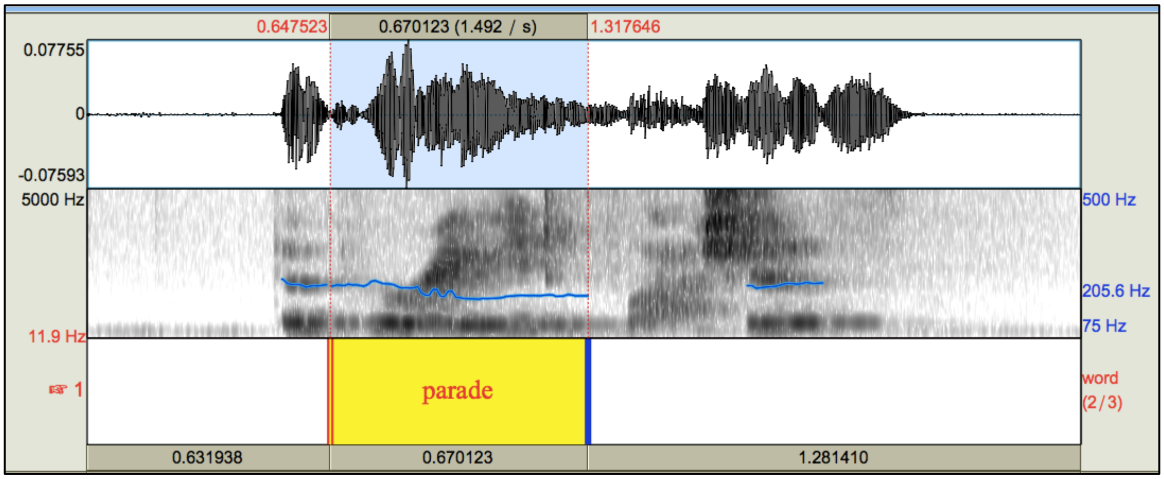


Figure 3. Example of Praat segmentation in Experiment 1. The .wav file is a recording of a participant saying, “The parade flashes,” and the Praat window shows waveform, spectrogram, and annotation of highlighted target word parade.

2.1.6 Inter-Rater Reliability

Inter-rater reliability was assessed by comparing a random subset of trials from all coders to a standard coder. This sample consisted of a reasonable number of trials (~10%) from each coder with equal sample size across coders. The intraclass correlation coefficient (ICC) was calculated using a one-way single-measures approach (Shrout & Fleiss, 1979; Hallgren, 2012) to determine agreement between coders. Experiment 1 had two coders who coded unique samples, and the author served as a standard coder who coded 36 trials from each of the coders' sets. Trials were randomly selected using a random number generator. The standard coder was blinded to the original measurements and experimental condition. The ICC was calculated between each coder and the standard (ICC_{Coder1} = 0.953, ICC_{Coder2} = 0.909), and the average of these was ICC = 0.931, indicating excellent agreement between coders.

2.2 RESULTS

Target word durations across conditions were analyzed, and only target utterances that matched the intended label were considered in the analyses. Trials were excluded if participants mispronounced the prime or target, or if they used alternate names (e.g., *boat* for *canoe*, *orchestra* for *quartet*, *cologne* for *perfume*). A total of 97 out of 1242 trials met these criteria and were removed. Scripts and the complete data set are available at https://osf.io/zk4qv/.

2.2.1 Effect of Condition on Word Duration

To examine the effects of condition on word duration, results were analyzed using a linear mixed effects model with a maximal random effects structure (Barr et al., 2013). That is, the model had condition as a fixed effect and random slopes and intercepts by item and by participant. Models were built using R package *lme4* version 1.1-10 (Bates et al., 2015). Data were log transformed and centered. Helmert contrasts were used in model development, such that each condition was compared with the average of its subsequent conditions. Significance was assumed for *t*-values with an absolute value greater than 1.96 in a two-tailed test (Baayen, 2008). This is an appropriate method in mixed-effects modeling because the number of degrees of freedom is large (> 100), so the *t*-distribution approximates the normal distribution.

We found that target items with segmental and metrical overlap were significantly longer than target items in the other conditions ($\beta = 0.047$, $t = 3.729$), and target items in the segmental overlap condition were significantly longer than target items with no overlap ($\beta = -0.033$, $t = -2.446$). Table 1 displays parameter estimates for the model. Additionally, iambs were significantly longer than trochees ($\beta = -0.138$, $t = -2.707$), regardless of condition; there was no interaction between meter type and overlap condition. Figure 3 displays average target durations by condition for this experiment.

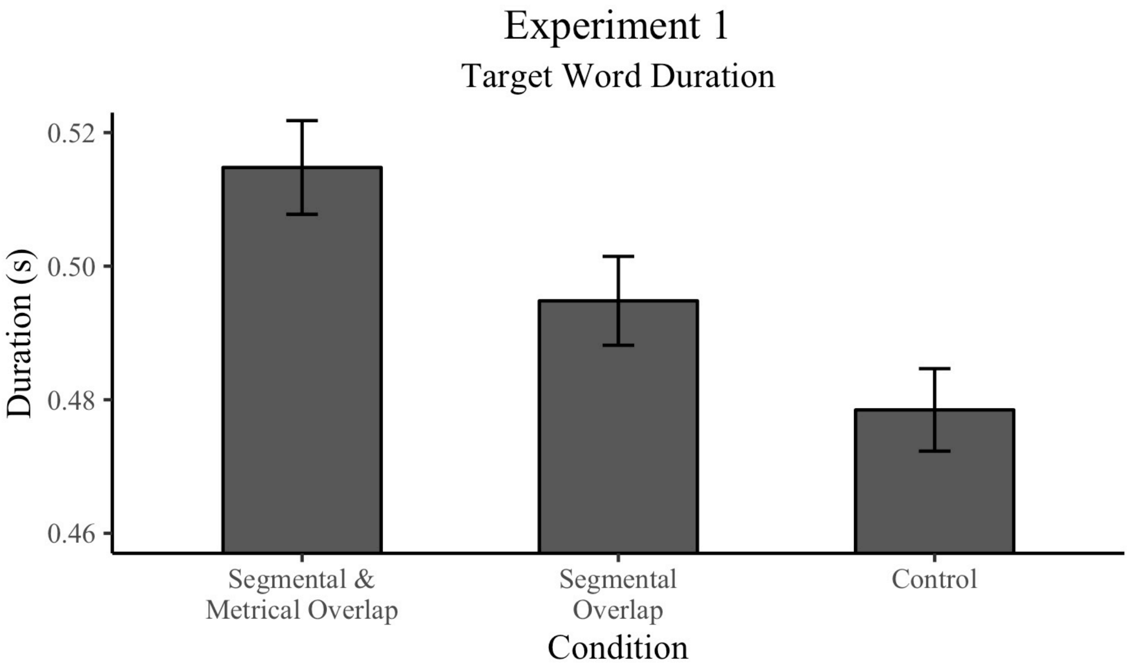


Figure 4. Results for Experiment 1. Average duration (in seconds) of target words by condition in Experiment 1. Error bars represent standard error for each condition.

Experiment 1

Fixed Effects	***β***	***t***
*Segmental & Metrical vs Segmental, Control**	0.047	3.729
*Segmental vs Control**	-0.033	-2.446

Table 1. Fixed effects estimates for Experiment 1. Fixed effects consisted of the Helmerted contrasts between conditions. Asterisks indicate significance.

2.2.2 Word Frequency Analysis

Word frequency refers to the occurrence of a word in a given text corpus, and frequency is associated with faster response times in picture naming (Jescheniak & Levelt, 1994). To confirm that word lengthening was not affected by frequency, we conducted a post hoc statistical test investigating whether word frequency contributed to duration. The frequency for all target words was captured from the SubtlexUS database (Brysbaert & New, 2009). The model described above was modified to include word frequency as a control variable, and these models did not differ significantly from each other ($\chi^2 = 1.54$, $p = 0.91$). Because word frequency does not improve the fit of the model, and we had no *a priori* predictions about word frequency, we did not consider it further.

2.2.3 Latent Semantic Analysis

Semantic interference has been shown to delay response times when naming pictures that are semantically related (Shao et al., 2013). To confirm that word lengthening was not affected by semantic information, we conducted a second follow-up statistical test using latent semantic analysis (LSA), which generates the degree of semantic similarity between two words. To determine semantic relatedness between words, we used a pairwise comparison application (lsa.colorado.edu) with a semantic space of college-level general reading, which yielded LSA scores from -1 to 1 for each word pair, where 0 means no similarity. The mean LSA score for word pairs in Experiment 1 was $M = 0.05$ ($SD = 0.07$). The model described above was modified to include semantic similarity as a control variable, and these models did not differ significantly from each other ($\chi^2 = 3.03$, $p = 0.70$). Because LSA does not improve the fit of the model, and the study was not designed to examine semantic relatedness, we did not consider LSA further.

2.3 DISCUSSION: EXPERIMENT 1

In this experiment we replicated previous findings that have shown that segmental overlap leads to significant word lengthening compared to prime-target pairs that do not overlap. We also found that the addition of metrical overlap leads to even more lengthening. This is potentially consistent with the notion that segmental and metrical spell-out occur through separate but similar processes (e.g., Roelofs & Meyer, 1998). However, it is unclear whether this effect is due to overlapping metrical structure or due to surface-level acoustic similarity between words. That is, the overlapping syllable in words with the same stress pattern (*candy/candle*) sound more alike than overlapping syllables in words with a conflicting stress pattern (*canteen/candle*). Thus, it is possible that surface-level acoustic similarity is driving the lengthening/competition effect.

This question is critical because it addresses a central theme of this dissertation: do segmental and metrical planning occur independently or do they share representations? Because overlapping metrical structures elicit longer target word durations, it is possible that stressed syllable locations are planned serially and induce lexical competition between words that have similar metrical structure. Conversely, it is possible that metrical overlap does not interfere with metrical planning, but rather, a representation that encodes acoustic-phonetic details of a word is planned serially and induces lexical competition between words that sound similar. These possibilities will be examined further in Experiment 3.

This experiment showed lengthening associated with segmental and metrical overlap, so it is possible that metrical values are encoded in serial order, much like the planning of phonemes (Sevald & Dell, 1994). In Experiment 2, we examine the sequential cueing of metrical information with overlapping initial segments. Sevald & Dell (1994) used a phrase repetition experiment that supported the serial ordering of phonemes, and Experiment 2 is designed with comparable methods to determine if meter is planned in a similarly serial manner.

CHAPTER 3: EXPERIMENT 2

Experiment 1 used an event description task and found word lengthening when words shared segmental and metrical information with recently produced words, which suggests that these features create competition during phonological encoding. Another approach to measuring this competition is by using a parameter-remapping paradigm (Rosenbaum et al., 1986), in which participants repeat sequences as many times as possible in a fixed time period (à la tongue twisters). Sevald and Dell (1994) conducted such a task where speakers repeated word pairs, and they found fewer repetitions when words had initial segmental overlap (e.g., *pick-pin*) compared to final segmental overlap (e.g., *pick-tick*). They concluded that when initial segments are the same, the activation of the initial segment creates competition for the next segment to be activated, and thus the sequence must be "remapped". If metrical values are planned similarly, then repetitions of the same stress pattern (e.g., *candle-candy*) should create more interference than opposing stress patterns (e.g., *candle-canoe*). In this experiment, we used a word pair repetition task (as in Sevald & Dell, 1994) and manipulated the location of stress in disyllabic words.

3.1 METHODS

3.1.1 Participants

Sixty native English speakers (age range: 18-32, *M* = 19.9 years, *SD* = 2.8, 47 female) were recruited to participate in this study. Recruitment procedures were the same as Experiment 1, and participants provided written informed consent in accordance with the Vanderbilt University Institutional Review Board. Two participants were excluded from analysis: one began but was unable to complete the experiment due to scheduling, and one had a speech sound disorder that inhibited accurate task performance.

3.1.2 Materials

A set of 80 disyllabic word pairs were used in this task, and all participants were exposed to the entire set. There were no filler items in this experiment. All word pairs shared initial segmental overlap and were arranged into four conditions based on the location of stressed syllables:

1. **trochee-trochee** (e.g., *ballot—ballad*)
2. **iamb-iamb** (e.g., *ballet—balloon*)
3. **trochee-iamb** (e.g., *ballot—ballet*)
4. **iamb-trochee** (e.g., *ballet—ballot*)

Conditions 1 and 2 had the same meter in both words, and conditions 3 and 4 had words with opposing meter. For the different meter conditions, we included trials that began with a trochee (condition 3) and trials that began with an iamb (condition 4). Iambic words tend to have longer durations and are more error-prone than trochaic words (Aichert et al., 2016), so both orders were used to measure an effect of initial stress. See Appendix C for the complete list of stimuli.

3.1.3 Procedure

Participants completed the experiment on a Mac computer in Matlab using the CogToolbox (Fraundorf et al., 2014) and Psychophysics Toolbox 3 (Kleiner et al., 2007).

Participants were instructed that they would see a two-word phrase on the computer, and their task was to repeat the phrase as fast as possible continuously for eight seconds. The experiment began with three practice trials. For each trial, participants saw the word pair printed in the middle of the computer screen for four seconds. Then, the words disappeared and the word “GO!” appeared. Participants then recited the word pair aloud continuously as fast as possible. After eight seconds of repetitions, an audible tone sounded and the word “STOP” appeared on the screen. Participants initiated the next trial by pressing the space bar. The experiment consisted of 3 practice trials and 80 test trials. All trials were randomized for each participant, and participants were allowed to take short breaks between trials as needed.

3.1.4 Acoustic Analysis

As in Experiment 1, speech recordings were analyzed in Praat (Boersma & Weenink, 2014) using manual segmentation to code the start and end times of phrase repetitions. Coders ran a Praat script that displayed each 8-second audio recording with spectrographic and waveform information, along with an annotation tier. Coders were trained on segmenting the phrases within the recorded utterance; segments consisted of both words in the target phrase. As the recording consisted of numerous phrase repetitions, coders zoomed in as needed to get a clear view of the spectrogram and waveform for one repetition. They used both audio and visual cues to demarcate the start and end points of the full phrase so that both words in the phrase exist between the marked boundaries. Coders were trained as in Experiment 1 regarding phoneme identification based on spectrographic information. They placed a boundary at the start of the initial phoneme of the first word and a boundary at the end of the final phoneme of the second word. They labeled the phrase by typing both words in between the boundaries on the annotation tier. If the 8-second recording stopped before a participant could finish a phrase, the coders did not segment this final incomplete utterance. When participants made speech errors, those segments were labeled as errors. Coders played back the phrase in isolation to determine if they accurately identified the full length of the phrase and nothing else. Then they adjusted the boundaries to ensure accurate segmentation. An example of this segmentation is provided in Figure 3. They continued segmenting all repetitions in a trial, and the Praat GUI then saved their annotation as a .TextGrid file and opened the next trial. See Appendix B for the coding manual that was provided to research assistants.

The duration for each phrase repetition was extracted from all trials using a Praat script that captured the amount of time between all annotated segments. This script provided output signifying participant, target phrase, number of repetitions, duration of each repetition, and number of errors for each trial, and this dataset was used in statistical analysis.

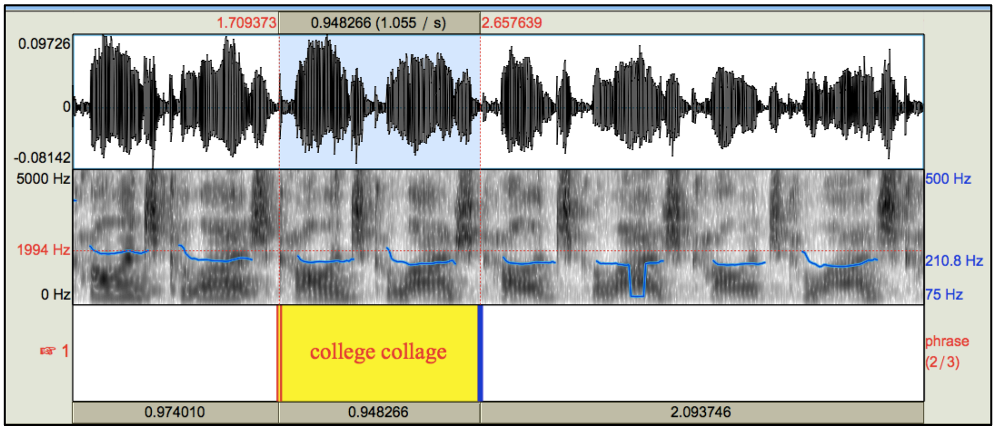


Figure 5. Example of Praat segmentation in Experiment 2. The .wav file is a recording of a participant repeating the phrase, "college collage," and the Praat window shows waveform, spectrogram, and annotation of the phrase. For clarity, one repetition is highlighted in this figure, but coders annotated each repetition in the recording.

Additionally, as a post hoc exploratory analysis, coders were trained to descriptively classify speech errors. To do this, they ran a Praat script that opened each segment that was previously annotated as an error. The script opened the spectrographic and waveform window at the start and stop boundaries of the error, along with 0.75 seconds on either end. Coders were given the target phrase, and they compared this to the actual utterance to determine the manner of speech error. They transcribed the participant's error and used the following classifications to identify error type:

1. Substitution (movement of phonemes that exist in the sequence)
2. Insertion (addition of phonemes not formally in the sequence)
3. Deletion (removal of phonemes without substitution)
4. Word repetition (production of entire word twice in a row)
5. Word shift (reversal of word order in sequence)
6. Vowel distortion (incorrect production of vowel or blend of vowels)

Coders were trained by completing 40 trials with the first author to establish agreement. If a participant consistently said the wrong word throughout a trial, that trial was excluded from analysis. This was considered an incorrect response rather than a series of speech errors. See Appendix D for the speech error coding manual.

3.2.5 Inter-Rater Reliability

Inter-rater reliability was assessed by comparing a random subset of duration measurements from all coders to a standard coder. This sample consisted of ~10% from each coder with equal sample size across coders. The intraclass correlation coefficient (ICC) was calculated using a one-way single-measures approach (Shrout & Fleiss, 1979; Hallgren, 2012) to determine agreement between coders. Experiment 2 had three coders who coded unique samples, and the author served as a standard coder who coded 32 trials from each of the coders' sets. Trials were randomly selected using a random number generator. The standard coder was blinded to the original measurements. The ICC was calculated between each coder and the standard ($\text{ICC}_{\text{Coder1}} = 0.979$, $\text{ICC}_{\text{Coder2}} = 0.925$, $\text{ICC}_{\text{Coder3}} = 0.944$), and the average of these was ICC = 0.949, indicating excellent agreement between coders.

3.2 RESULTS

Phrase durations across conditions were analyzed, and only target utterances that matched the intended phrase were considered in the analyses. Trials were excluded if participants consistently mispronounced the target phrase throughout the trial (e.g., *despot* pronounced as [ˈdɛspoʊ], *desert* pronounced as [dɪˈzɜrt]). A total of 54 out of

4800 trials met these criteria and were removed. Scripts and the complete data set are available at https://osf.io/zk4qv/. Additionally, this experiment was pre-registered with the Open Science Framework (Myers & Watson, 2018).

3.2.1 Effect of Meter on Phrase Duration

To examine the effects of metrical condition on phrase duration, results were analyzed using a linear mixed effects model with a maximal random effects structure (Barr et al., 2013). A model with this structure informs whether there is a main effect of overlapping metrical structure, a main effect of initial syllable stress, and an interaction effect between these two variables. Because the maximal model did not converge, specification of the random effects structure was systematically simplified by varying random slopes and intercepts for within-unit (participant or item) factors. Results presented are from the best fitting model based on AIC, BIC, and log-likelihood scores. The model was built using R package *lme4* version 1.1-10 (Bates et al., 2015). Data was log transformed and centered. Significance was assumed for *t*-values with an absolute value above 1.96 in a two-tailed test as in Experiment 1 (Baayen, 2008).

There was a significant main effect of metrical overlap on average phrase duration ($\beta = -0.039$, $t = -2.468$), but there was no effect of initial stress, nor was there an interaction. That is, word pairs with matching meter had significantly longer durations than word pairs with different meter, regardless of the stress of the initial syllable. Similarly, there was a significant main effect of metrical overlap on the number of repetitions per trial ($\beta = 0.087$, $t = 3.878$), but there was no effect of initial stress, nor an effect of their interaction. Figure 4 displays average phrase durations and number of repetitions across conditions, and Table 2 displays parameter estimates for the models.

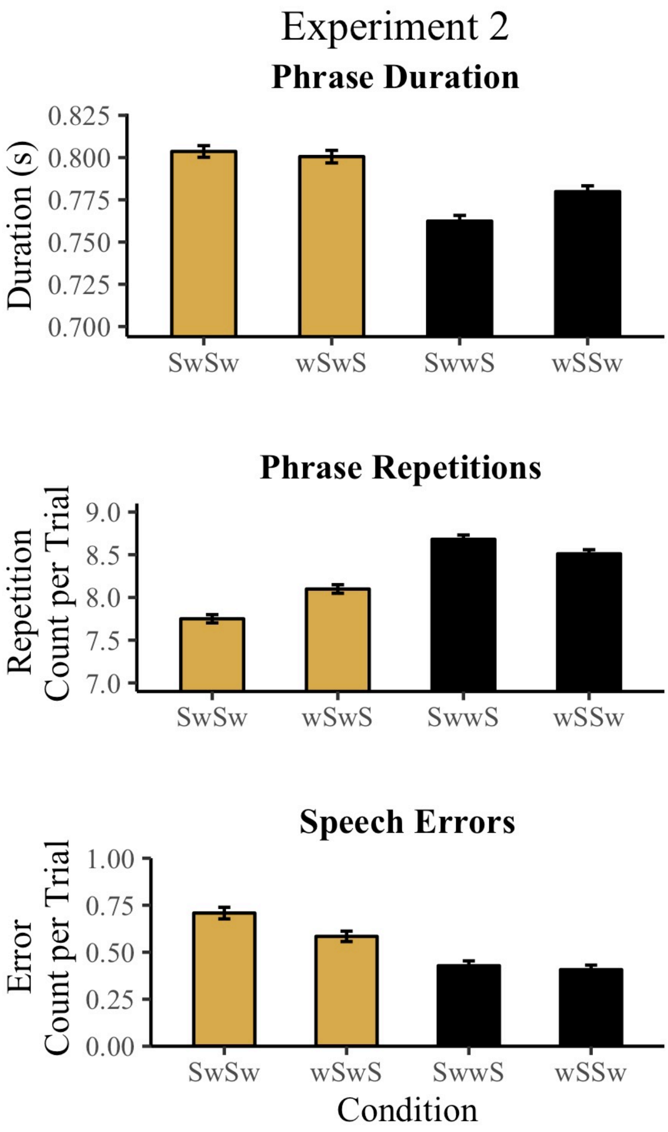


Figure 6. Results for Experiment 2. Top panel: average phrase duration (length of time to produce both words) in seconds. Middle panel: average number of phrase repetitions per trial. Bottom panel: average number of speech errors made per trial. Gold bars indicate overlapping metrical structure; black bars indicate different metrical structure. Error bars represent standard errors.

Experiment 2

Fixed Effect: Metrical Overlap	***β***	***t***
*Duration**	-0.039	-2.468
*Repetitions**	0.087	3.878
*Speech Errors**	-3.13	-5.405
Initial Stress	***β***	***t***
Duration	0.008	0.510
Repetitions	0.014	0.514
Speech Errors	-0.617	-1.103
Interaction	***β***	***t***
Duration	0.030	0.933
Repetitions	-0.067	-1.535
Speech Errors	1.199	1.039

Table 2. Fixed effects estimates for Experiment 2. Estimates are for metrical overlap in phrase duration, number of repetitions, and number of speech errors. Asterisks indicate significant findings. No significant findings were observed for the fixed effect of initial stress, or for the interaction between initial stress and metrical overlap.

3.2.2 Speech Errors

The number of speech errors was counted for each trial as an indicator of when remapping failed. There was a significant main effect of metrical overlap on the number of speech errors produced ($\beta = -3.13$, $t = -5.405$), but there was no effect of initial stress, nor an effect of their interaction. Additionally, as a post hoc exploratory analysis we descriptively categorized speech errors to determine the types of errors that participants made during the task. Coders classified a total of 2345 speech errors into six distinct error categories (described in section 3.1.4). Overall, the conditions with metrical overlap had more speech errors ($n = 1421$) than the conditions with different meter ($n = 924$). Figure 5 is a graphical display of speech errors by metrical overlap condition, and Table 3 provides the inventory of error types by counts and percentages. A chi-square test of independence revealed that metrical conditions had distinct distributions of speech error types ($\chi^2 = 26.4$, $p < 0.01$). Post hoc analyses revealed that there were more word repetitions in same meter conditions ($\chi^2 = 10.3$, $p < 0.01$), and there were more phoneme deletions in different meter conditions ($\chi^2 = 3.8$, $p = 0.01$).

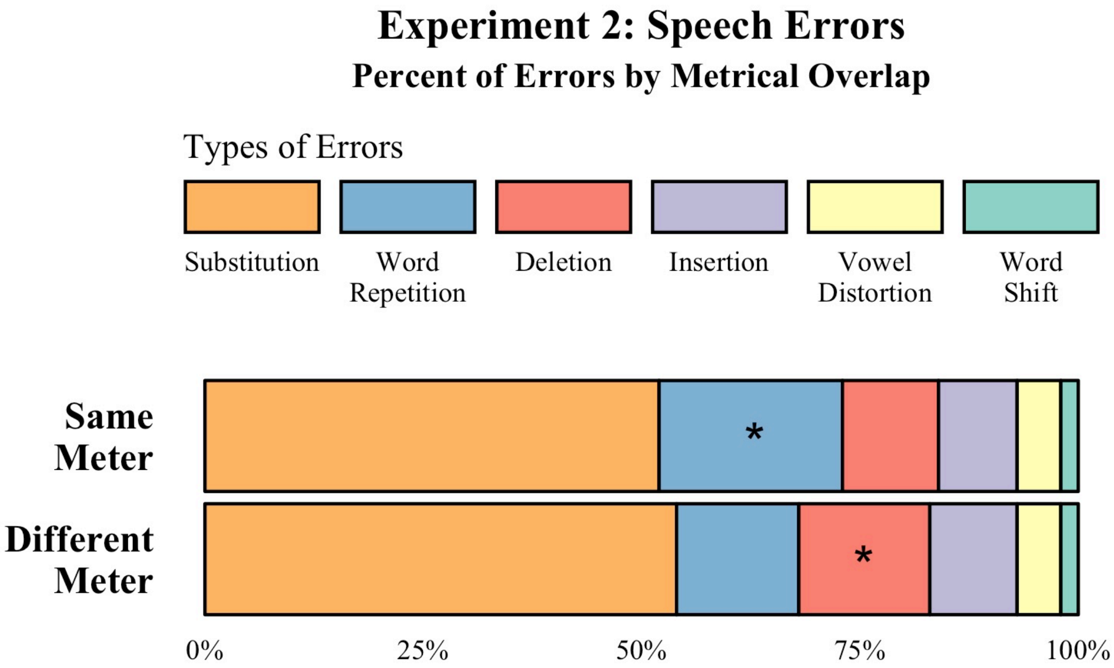


Figure 7. Distribution of speech errors in Experiment 2. Stacked bars represent percentage of errors for each condition. Percentages are based on a total of 1421 errors for Same Meter trials and 924 errors for Different Meter trials. Error categories that are proportionally distinct across metrical conditions are indicated with an asterisk.

Experiment 2

Error Type	**Same Meter**	**Different Meter**	**Total**
Substitution	740 (52%)	497 (54%)	1237
Word Repetition	303 (21%)	128 (14%)	431
Deletion	163 (11%)	141 (15%)	304
Insertion	124 (9%)	91 (10%)	215
Vowel Distortion	74 (5%)	49 (5%)	123
Word Shift	17 (1%)	18 (2%)	35
Total	1421	924	2345

Table 3. Speech errors in Experiment 2. Number of speech errors in each error category by metrical condition. Percentages reflect proportion of errors within metrical condition.

3.2.3 Word Frequency Analysis

To confirm that phrase lengthening was not driven by an effect of word frequency, we conducted a post hoc statistical test investigating whether word frequency contributed to the effect of metrical condition on phrase duration. Word frequency refers to the occurrence of a word in a given text corpus, and the corpus used for this study was the SubtlexUS database (Brysbaert & New, 2009). The frequency for all target words was captured, and the frequency for each word pair was calculated by the sum of the two frequencies in the pair. The sum of frequencies was used because the word pairs did not have any occurrences as a single unit in the database; for instance, *beaker beagle* never appeared as a phrase in the database, so we added the frequency of *beaker* and the frequency of *beagle* for the trial. The models described above did not differ significantly when including word frequency as a control variable for phrase duration ($\chi^2 = 1.88$, $p = 0.17$), repetitions ($\chi^2 = 8.09$, $p = 0.32$), and speech errors ($\chi^2 = 10.41$, $p = 0.11$). Because word frequency does not improve the fit of the models, and we had no *a priori* predictions about word frequency, we did not consider it further.

3.2.4 Latent Semantic Analysis

To confirm that phrase lengthening was not driven by an effect of semantic information, we conducted a second follow-up statistical test using latent semantic analysis (LSA), which generates the degree of semantic similarity between two words. To determine semantic relatedness between words, we used a pairwise comparison application (lsa.colorado.edu) with a semantic space of college-level general reading, which yielded LSA scores from -1 to 1 for each word pair, where 0 means no similarity. The mean LSA score for word pairs in Experiment 2 was $M = 0.10$ ($SD = 0.12$). The models described above did not differ significantly when including LSA as a control variable for phrase duration ($\chi^2 = 7.34$, $p = 0.29$), repetitions ($\chi^2 = 7.25$, $p = 0.40$), and speech errors ($\chi^2 = 3.05$, $p = 0.80$). Because LSA does not improve the fit of the models, and the study was not designed to examine semantic relatedness, we did not consider LSA further.

3.3 DISCUSSION: EXPERIMENT 2

3.3.1 Effect of Metrical Overlap

In this experiment word pairs in the four conditions were organized by same or different stress patterns and a stressed or unstressed initial syllable. We find evidence that metrical overlap in words leads to phrase lengthening, fewer repetitions, and more speech errors. This finding is very similar to previous work showing slowed productions with phonemic overlap (e.g., Sevald & Dell, 1994), but this is the first time—to our knowledge—showing the effect in meter. The sequential cueing model suggests that when target phonemes are activated in a word, they also activate other words that share those phonemes, which causes sequencing difficulty in the system (Sevald & Dell, 1994; Meyer, 1991). Now we see that meter also plays a role in the speed of word production. When meter is consistent across words with overlapping phonemes, the system requires additional time to sort out the sequential cueing of the target words. It may be possible that metrical encoding occurs on a separate tier than segmental encoding (as in Roelofs & Meyer, 1998; Keating & Shattuck-Hufnagel, 2002). If that were the case, then in a parameter-remapping paradigm such as this one, the metrical structure would not need to be remapped because it stays consistent on its separate tier. In that scenario, the established metrical structure should expedite word production, much like the effect when repeated words have shorter durations (e.g., Kahn & Arnold, 2015). However, Experiment 2 shows the opposite effect, such that activating a consistent metrical structure is indeed more challenging than using varied stress patterns, which indicates that meter may not be established as its own parameter.

A more likely explanation for why this occurs may be that meter is not planned as a separate entity from phonemes, but rather the two representations may be merged during the planning process. Because we have reason to believe that meter does not exist independently (above), it is plausible that the competition effect is driven by surface level acoustic properties of the words (as discussed in Experiment 1). For example, in a phrase like *corner coral*, the initial segment of both words (*cor-*) is more or less identical in phonemes and stress. Whereas the phrase *corner correct* has a similar initial sequence of phonemes (*cor-*), but the initial stress is different, making the first syllables acoustically distinct. Because *corner coral* elicits more competition and production difficulty, it is possible that the observed word lengthening is determined by the overall acoustic similarity between the words. To adjudicate between these two possible explanations (representation of abstract metrical structure versus acoustic-phonetic details), we conducted Experiment 3 to investigate word lengthening in word pairs that shared metrical structure without phonemic overlap.

3.3.2 Speech Error Analysis

Our investigation into trends of speech errors was conducted as a speculative post hoc treatment, and we did not make predictions about this behavior in our research question. That being said, the sequential cueing model (Sevald & Dell, 1994) asserts that phonological encoding occurs from left to right, and competition for shared segments can incorrectly cue sounds later in the sequence. This was justified in our observation of phoneme substitutions as the largest category of speech errors; that is, the most common error was due to phonemes being cued out of order in the sequence. The number of substitutions was relatively consistent regardless of whether word pairs had the same or different metrical structure. The sequential cueing model predicts substitution errors between words that share the same initial phonological segments, and all of the word pairs in this experiment fit that description. Therefore, we find support for the model, but we are not able to draw conclusions about how meter may fit into that model because these are still speculative findings.

Word repetitions were the second most prominent error category, and word pairs with the same metrical structure had significantly more word repetitions than different meter pairs. Word repetitions are much like substitutions in that they both rearrange phonemes for a new order in the sequence, only word repetitions miscue an entire word. Interestingly, this full word miscuing occurs more frequently in word pairs with matching meter, which indicates that meter does play a role in the cueing process. As discussed above, word pairs with the same initial segment and stress have greater acoustic similarity than words with opposing stress, and this potentially creates greater interference when cueing the next word in a sequence.

Phoneme deletions were observed proportionally more in word pairs with different metrical structure, and this is likely an indirect effect of phonological encoding. Deletions were characterized as simply removing a phoneme from the sequence without replacing it with another phoneme. Word pairs with alternating stress patterns were produced with a faster rate, and as with any rapid motor task, accuracy in achieving targets is often compromised. Therefore, deletions were not necessarily a reflection of the phonological plan and more so a consequence of motor performance.

Additionally, we excluded trials where words were incorrectly pronounced throughout the trial. These incorrect trials occurred more often in word pairs with the same meter (n = 46) than word pairs with different meter (n = 8). The same meter targets were sometimes incorrect because speakers would change the stress of the second word to create different metrical structures. For example, the same meter phrases *corner coral* and *secure secrete* were sometimes produced with different meter as *corner chorale* and *secure secret*. This desire to change the metrical structure so that the words are not the same could indicate that speakers are aware (at some level) that the same meter pairs are more difficult to produce, so they devise an alternate plan to make the task easier. It is possible that speakers simply misread the target words, but the fact that this happened more in the same meter conditions alludes to a performance challenge in these trials. Even though this occurred in a very small fraction of trials, it raises the question again of how independent meter is from segmental spell-out. To specifically target this issue of metrical independence, we introduce Experiment 3, which focuses on distinguishing segmental and metrical spell-out in phonological encoding.

CHAPTER 4: EXPERIMENT 3

The third experiment was designed to determine the independence of metrical spell-out from segmental spell-out in phonological encoding. While Experiments 1 and 2 observed additional word lengthening with segmental and metrical overlap, this does not test the assumption that segmental and metrical spell-out are separate processes (as in Roelofs & Meyer, 1998). In Experiment 3, we used the same task from Experiment 1 and introduced a condition specifically testing metrical overlap alone to examine whether meter is indeed planned independently. If meter is planned as an independent process—as predicted in the WEAVER model—we expect to observe word lengthening when words share only the same stress pattern because there should be competition for retrieving the initial stress value. If meter is not independent of segments, we expect no word lengthening in the metrical overlap alone condition because the competition for word retrieval is minimal. The lack of lengthening would be consistent with the acoustic similarity hypothesis from the previous experiments and suggest that meter is coupled with phonemic segments during phonological encoding.

4.1 METHODS

4.1.1 Participants

Sixty native English speakers (age range: 18-32, *M* = 19.9 years, *SD* = 2.8, 47 female) participated in this study. Recruitment procedures were the same as Experiment 1, with the caveat that participants were not permitted to participate in both experiments.

One participant was excluded from statistical analysis due to technical difficulties in recording.

4.1.2 Materials

A set of 160 color images was selected from the Snodgrass and Vanderwart (1980) dataset (Rossion & Pourtois, 2001) and Clipart, which included 80 critical items and 80 filler items. Critical items consisted of 16 targets and 64 primes. Prime-target pairs were arranged into four conditions:

1. **Metrical & segmental overlap:** The *ballet* shrinks. The *balloon* flashes.
2. **Segmental overlap alone:** The *ballot* shrinks. The *balloon* flashes.
3. **Metrical overlap alone:** The *guitar* shrinks. The *balloon* flashes.
4. **No overlap:** The *trumpet* shrinks. The *balloon* flashes.

We used a Latin square design, which yielded four counterbalanced lists of items, such that each participant was exposed to 16 critical prime-target pairs—four pairs for each of the four conditions. An equal number of trochees and iambs were used in each list. In addition, participants were exposed to 32 non-critical pairs, drawn from the filler items, for a total of 48 trials in the experiment. See Appendix E for the list of critical prime-target pairs.

4.1.3 Procedure

The design and instructions were the same as in Experiment 1. The primary difference between experiments was in the materials used. The predictions and data analysis strategy for Experiment 2 were pre-registered through the Open Science Framework (Myers & Watson, 2018).

4.1.4 Acoustic Analysis

Acoustic analysis was performed in the same manner as Experiment 1. Speech recordings were analyzed in Praat (Boersma & Weenink, 2017), using manual segmentation to code the start and end times of target words. Coders were blind to the experimental condition of trials. Target words were segmented such that they were not identifiable as anything other than the targets. Coders were trained on segmentation as in Experiments 1 and 2. As in Experiment 1, the duration for each target word was extracted from all trials using a Praat script that captured the amount of time between the two boundaries on either side of the target word. This script provided output signifying participant, condition, target word, and duration for each trial, and this dataset was used in statistical analysis. See Appendix B for the coding manual used in data analysis.

4.1.5 Inter-Rater Reliability

Inter-rater reliability was assessed by comparing a random subset of trials from all coders to a standard coder. This sample consisted of a reasonable number of trials (~10%) from each coder with equal sample size across coders. The intraclass correlation coefficient (ICC) was calculated using a one-way single-measures approach (Shrout & Fleiss, 1979; Hallgren, 2012) to determine agreement between coders. Experiment 3 had two coders who coded unique samples, and the author served as a standard coder who coded 32 trials from each of the coders' sets. Trials were randomly selected using a random number generator. The standard coder was blinded to the original measurements and experimental condition. The ICC was calculated between each coder and the standard (ICC_{Coder1} = 0.948, ICC_{Coder2} = 0.958), and the average of these was ICC = 0.953, indicating excellent agreement between coders.

4.2 RESULTS

Target word durations across conditions were analyzed, and only target utterances that matched the intended label were considered in the analyses. Trials were excluded if participants mispronounced the prime or target, or if they used alternate names (e.g., *boat* for *canoe*, *orchestra* for *quartet*, *cologne* for *perfume*). A total of 59 out of 944 trials met these criteria and were removed. Scripts and the complete data set are available at https://osf.io/zk4qv/.

4.2.1 Effects of Segmental and Metrical Overlap on Duration

Maximal mixed effects models were built in the same manner as in Experiment 1 to examine the fixed effects of segmental overlap, metrical overlap, and their interaction. An ANOVA was carried out to determine the best fitting model, which included random slopes and intercepts for the segmental by metrical interaction by item, as well as the segmental and metrical manipulations by participant.[1] Data were log-transformed and

---

[1] Two models converged that were one step down from the maximal model. These models did not differ significantly from one another. Both models yielded a significant segmental by metrical overlap interaction using the 1.96 criteria. However, the *lmerTest* function yielded an interaction that was marginal for one model but significant for the other. We report findings from the best fitting model based on AIC, BIC, and log-likelihood scores although the interaction in this model (t=-2.030) was marginally significant according to *lmerTEST* but significant using the 1.96 threshold.

centered. Significance was assumed by *t*-values with absolute value above 1.96 in a two-tailed test as in Experiment 1 (Baayen, 2008). There was a significant main effect of segmental overlap ($\beta = 0.054$, $t = 5.038$) and a significant interaction between segmental and metrical overlap ($\beta = -0.038$, $t = -2.030$). No main effect of metrical overlap was observed. Table 2 displays parameter estimates for the model. Additionally, iambs had longer durations than trochees on average ($\beta = -0.162$, $t = -3.216$), regardless of condition; there was no interaction between metrical type and overlap condition. Figure 6 displays average target durations by condition for this experiment, and Table 3 displays model statistics. Appendix F summarizes findings from all three experiments.

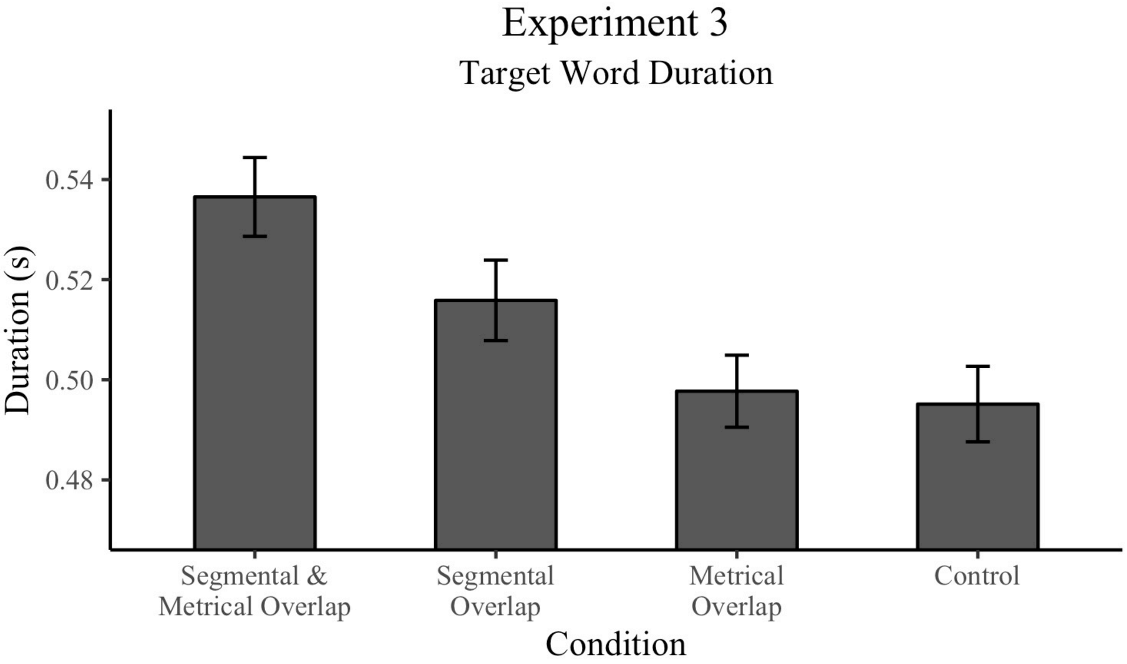


Figure 8. Results for Experiment 3. Average duration (in seconds) of target words by condition. Error bars represent standard error for each condition.

Experiment 3

Fixed Effects	***β***	***t***
*Segmental Overlap**	0.054	5.038
Metrical Overlap	-0.017	-1.714
*Segmental x Metrical**	-0.038	-2.030

Table 4. Fixed effects estimates for Experiment 3.

4.2.2 Word Frequency Analysis

To confirm that word lengthening was not driven by an effect of word frequency, we conducted a post hoc follow-up statistical test investigating whether word frequency contributed to the effects of segmental and metrical overlap on word duration. Word frequency refers to the occurrence of a word in a given text corpus, and the corpus used for this study was the SubtlexUS database (Brysbaert & New, 2009). The frequency for all target words was captured. The model described above was modified to include word frequency as a control variable, and these models did not differ significantly from each other ($\chi^2 = 2.89$, $p = 0.89$). Because word frequency does not improve the fit of the model, and we had no *a priori* predictions about word frequency, we did not consider it further.

4.2.3 Latent Semantic Analysis

To confirm that word lengthening was not affected by semantic information, we conducted a second follow-up statistical test using latent semantic analysis (LSA), which generates the degree of semantic similarity between two words. To determine semantic relatedness between words, we used a pairwise comparison application (lsa.colorado.edu) with a semantic space of college-level general reading, which yielded LSA scores from -1 to 1 for each word pair, where 0 means no similarity. The mean LSA scores for word pairs in Experiment 3 was $M = 0.05$ ($SD = 0.07$). The model described above was modified to include semantic similarity as a control variable, and these models did not differ significantly from each other ($\chi^2 = 3.00$, $p = 0.88$). Because LSA does not improve the fit of the model, and the study was not designed to examine semantic relatedness, we did not consider LSA further.

4.3 DISCUSSION: EXPERIMENT 3

We replicated the findings from Experiment 1 by showing that segmental overlap alone leads to significant word lengthening, and metrical and segmental overlap lead to even longer word durations. However, speakers did not lengthen words with metrical overlap alone; word durations in this condition were no different from those in the control condition without any overlap. Although we hesitate to draw conclusions from a null result, the data are most consistent with lengthening being contingent on surface-level acoustic similarity of words. That is, word pairs with overlapping segments and meter (e.g., *locket* and *locker*) have more acoustic similarity than word pairs with opposing meter (e.g., *locket* and *lacrosse*), which have more acoustic similarity than word pairs with only meter in common (e.g., *locket* and *spider*). The level of acoustic-phonetic likeness in our word pairs is reflected in the relative amount of word lengthening across conditions. Because meter by itself did not have a significant impact on word duration in Experiment 3, we do not see evidence of a planning stage for abstract metrical structure, but instead we argue that stress is closely bound to segmental representations during phonological encoding.

CHAPTER 5: GENERAL DISCUSSION

In three experiments, we tested the hypothesis that metrical and segmental spell-out occur as distinct but parallel processes (Roelofs & Meyer, 1998). We used word lengthening driven by segmental and metrical overlap as an index of whether these two types of linguistic structure are planned independently or together. All three experiments revealed lengthening when word pairs had metrical and segmental overlap. Experiments 1 and 3 also showed lengthening to a lesser degree when primes and targets had segmental overlap alone. However, in Experiment 3 we found no evidence of lengthening from metrical overlap alone. This suggests that representations for metrical stress are tightly linked to segmental structure in speech production. We also conducted post hoc analyses on word frequency and semantic similarity to confirm that these processes were not contributing to the word lengthening effect. Furthermore, because segmental overlap leads to some degree of word lengthening regardless of metrical structure, it seems that phonemic representations play a more central role in word planning than metrical representations. These findings lead us to conclude that the word lengthening observed in these experiments may be driven by the acoustic representations of target words.

5.1 IMPLICATIONS AND FUTURE DIRECTIONS

5.1.1 Acoustic Representations in Phonological Encoding

One explanation for the lengthening observed in these experiments is that the surface-level, fine-grained acoustic properties of sounds—rather than abstract metrical and segmental representations—serve as the building blocks of speech planning. It is possible that the production system maintains individuated representations for stressed (/ca-/ in *candle*) and unstressed (/ca-/ in *canoe*) syllables and uses these bound representations when ordering the sounds of the word. There is ample evidence suggesting that contextual information is maintained in long-term linguistic representations alongside segmental information; this can include variability in productions, as well as features pertaining to gender, age, and dialect (see Pierrehumbert, 2016 for review). This suggests that at the very least, speakers maintain detailed acoustic forms of words and syllables, and this almost certainly would include whether a syllable is stressed or not.

There is also evidence that this type of detailed acoustic representation may play a pivotal role in the encoding process by serving as the basis of feedback to the production system during articulation. We know that auditory feedback is important in monitoring the state of the production system during articulation, and it has been argued that one function of auditory feedback is to provide information about deviations from the intended output (see Guenther, 2014; Hickok, 2014 for review). Data from Jacobs et al. (2015) suggests that phonological overlap interference effects may be driven in part by auditory feedback. They found that producing a prime aloud affects a subsequent target word's duration, but producing the prime as inner speech or silent mouthing does not. That is, the interference does not come from the lexical-conceptual or articulatory levels. Similarly, Buxó-Lugo et al. (in revision) found that subjects lengthened a target word when they produced a phonologically overlapping prime, and lengthening also occurred when a different speaker produced the prime. No target lengthening occurred when the subject silently mouthed the prime. These previous results show that the acoustic realization of a prime creates interference, possibly by influencing representations that are used in feedback mechanisms for speech encoding. Taken together, these studies suggest that detailed fine-grained acoustic representations of words and syllables may serve as building blocks for sequencing the sounds of words, possibly through auditory feedback. The results of the three experiments presented here suggest that these building blocks may consist of a detailed, unitary representation of segmental and metrical information.

5.1.2 An Integrated Model of Segmental and Metrical Planning

Our primary purpose for these experiments was to test the hypothesis that segmental and metrical spell-out occur as distinct but parallel processes (Roelofs & Meyer, 1998). When spoken words share initial phonemic segments, the word duration becomes longer (i.e., Yiu & Watson, 2015). Therefore, if meter is planned on a separate analogous track—as in the WEAVER model—then one might expect to see word lengthening when words share stress patterns. In Experiments 1 and 2, we indeed observed word lengthening related to metrical overlap (only in the presence of segmental overlap), and on the surface this may seem like support for the separate track model. That is, words lengthen with segmental overlap, and they lengthen even more by adding metrical overlap, so it could be possible that these are separate entities that contribute to word lengthening independently. However, Experiment 3 revealed that word lengthening does not occur from metrical overlap alone, which we suggest negates the possibility of an independent track for metrical spell-out. One could argue for the separate track model by suggesting that word lengthening requires segmental overlap first, and then metrical overlap creates a secondary lengthening effect. While this is theoretically plausible, a brief look into the speech motor planning literature will demonstrate that this is computationally inefficient and uncharacteristic of the speech production system (e.g., Guenther, 1995), and this is considered below.

It has been suggested that sequences of phonemes and sequences of syllables in common words and phrases are coded as larger units that can be retrieved as higher-level motor chunks (Cholin et al., 2006; Levelt & Wheeldon, 1994; Levelt et al., 1999). These established motor sequences allow for expedient production, and they may be stored in what Levelt calls a mental syllabary (Levelt, 1989). If this system of chunking were not in place, word representations would consist of individual phonemes, their sequence, as well as the location of stress. However, if a word is coded at a more holistic level, then phonemes and stress can be part of a higher-level code that efficiently guides motor control. Hickok (2014) suggests that acoustic targets are similarly chunked for motor sequences, so it is reasonable that acoustic stress would have a role in realizing these targets. This brings us back to the acoustic representations in speech planning. If the /ca/ of *candle* is different from that of *canoe*, it is conceivable that two representations of the syllable—stressed and unstressed—are stored in the syllabary. When a speaker retrieves this syllable, he or she is selecting the acoustic target with the appropriate stress value attached.

In this model, not all syllables will have stressed and unstressed versions readily available in the syllabary. The system must learn the relation between stress and phonemic segments, and only those sequences that are highly familiar will be “strongly chunked” in storage (Hickok, 2014). Take for instance the stressed syllable *chair*; in

American English, there is typically no case where this phonemic sequence appears unstressed, so it is likely stored with a strong connection to a stress accent. Now consider a sequence like *gest*. This most commonly appears in stressed form, such as *gesture*, *congested*, *digest*, but it can also appear unstressed as in *gestation*. Because this phonemic sequence is more frequently associated with a stress accent, the stressed version will likely be a stronger chunk than the unstressed version. Nevertheless, both versions may exist simultaneously in the syllabary, as with the aforementioned stressed and unstressed versions of /ca/.

One could argue that the stressed and unstressed versions of syllables are phonemically distinct and therefore do not refer to the same segments. For example, *candle* is pronounced with a low front vowel [kæ], but the unstressed vowel in *canoe* is a more neutralized [kə]. In our model, this is explained by placing feature-specific phonemes at a later stage in the speech processing system, as described in Hickok (2014). Our working hypothesis is that syllables are retrieved or formed with at least partial contribution from auditory targets, and then the context-specific motor trajectories ([kæ] vs. [kə]) are defined from somatosensory targets. We posit that the word lengthening from these experiments is due to interference at the level of syllable retrieval, and the difference in vowel shape is executed at a later stage. This model is consistent with the proposal from Hickok (2014), only now with the addition of metrical assignment as an auditory target in speech processing.

5.1.3 Implications for Rhythm and the Brain

Metrical regularity at the utterance level has been shown to facilitate language processing in speech perception. The ebbing and flowing stress patterns provide cues to the listener for parsing the unfolding speech signal (Pitt & Samuel, 1990), and an event-related potential (ERP) component has been observed related to processing metrical properties of speech, such that listeners are sensitive to when stress falls outside of a regular rhythm (Böcker et al., 1999). Even in the absence of a regular temporal rhythm, listeners generate metrical expectancies when listening to speech (Schmidt-Kassow & Kotz, 2009; Magne et al., 2016). Other ERP studies have shown that unexpected stress patterns lead to difficulties in lexical access (Friedrich et al., 2004; Knaus et al., 2007). By definition, stress contributes to rhythmic patterns, but how much is this rhythm dependent on segmental structures?

The acoustic make-up of a speech signal consists of fast-moving temporal fine structure (e.g., frequency characteristics of phonemes) and the subsequent temporal envelope (e.g., amplitude contour of syllables), which captures the broad variations in the signal. Speech information can be broken down to generic timescales where the envelope occurs at the rate of suprasegmental features of prosody, and the fine structure correlates with phonemic qualities (Keitel et al., 2018). Phonemes have a faster rate of production, which is nested within the slower rate of stressed syllables, and this hierarchy is reflected in neural entrainment to speech stimuli (Ding et al., 2017). Therefore, it is possible that phonemic segments may influence speech rhythm and likewise contribute to neural entrainment to the speech envelope (Di Liberto et al., 2015; see Myers et al., 2019 for review).

It has been suggested that English-learning infants have a preference to trochees rather than iambs (Jusczyk et al., 1993; Jusczyk, 1999), likely due to higher frequency in the language (Cutler & Carter, 1987). However, this preference appears in adult-directed speech but not infant-directed speech (Wang et al., 2016). Infant-directed speech reduces the salient phoneme-specific spectral cues, which provides fewer cues to contrast stressed and unstressed syllables. Therefore, perception of speech rhythm seems to be related to segmental structure. Indeed, linguistic theories have suggested that word stress is dependent on phonemic properties of a word (Halle & Vergnaud, 1987; Hayes, 1995); namely, elements that contrast in intensity form trochees, and elements that contrast in duration form iambs (*The Iambic/Trochaic Law*, Hayes, 1995). Iversen et al., (2008) showed that this rhythmic grouping is dependent on auditory experience, such that English-speaking adults could predict trochaic and iambic patterns, but Japanese-speaking adults could only predict trochaic sequences. Auditory experience influences rhythmic grouping and speech segmentation even in infancy (Hay & Saffran, 2012). Younger infants have been shown to rely more on phonemic cues, whereas older infants rely more on metrical cues (Thiessen & Saffran, 2003). This implies that listeners may develop metrical associations to phonemic properties with auditory experience. Therefore, it is likely that segmental and metrical structures become chunked together into a single auditory representation and contribute to the overall rhythm of speech, which is consistent with the findings from the current studies in this dissertation.

5.1.4 Future Directions

Additional experiments may be conducted to support the notion that segments and meter are combined into a singular auditory representation. In our studies, we measured durations of full words or phrases, but this does not allow us to determine which part of the word was particularly challenging. Watson et al. (2015) measured overlapping initial or final morphemes, and they found that speakers slowed down during the non-overlapping segment. This suggests that the production system provides more time for phoneme selection when faced with competition. A similar study could manipulate segmental and metrical structures to determine if they are separate or unified. For instance, metrical patterns could be consistent while changing the location of segmental overlap. As predicted by Yiu & Watson (2015), the non-overlapping segment should be longer in words with initial overlap (e.g., *candy-candle*) compared to final overlap (e.g. *girdle-candle*). Since we believe meter is tied to phonemic segments, the same results should also occur when stress is on the second syllable; that is, initially overlapping iambs (e.g., *canal-canoe*) should be longer than finally overlapping iambs (e.g., *renew-canoe*). Based on our findings herein, we believe that trochees and iambs will behave similarly to each other, but if the results vary based on stress pattern, that would contradict our hypothesis and indicate that meter uses unique mechanisms in speech production.

In the current experiments, our participants spoke aloud the prime-target word pairs, thus establishing motor chunks for phonemic segments. A subsequent experiment could control the acoustic properties of the chunked segments by having another speaker with a distinct dialect say the prime. That is, if the participant hears a segment that sounds different from how s/he would say it, lengthening may not occur because the perceptual acoustic cue does not interfere with the auditory target in production. This type of study design could be used to test dialects that vary in either phonemic features or metrical structures. In a separate experiment, pseudo-word primes could be used to test that this interference occurs at the level of lexical retrieval. Sevald and Dell (1994) predicted that words with pronounceable phonemic structure but no semantic meaning would not elicit word lengthening because they do not have an existing lexical representation.

The connection of meter and segments can also be tested in perceptual brain studies. Magne et al. (2016) showed sensitivity to speech meter by capturing an ERP response to words with unexpected stress patterns compared to others in a list. For instance, a trochee in a list of iambs (e.g., morale, embrace, delight, *pedal*) elicited an increased negativity response from the centro-frontal region of the scalp compared to the response elicited from a trochee in a list of trochees (e.g. zebra, bacon, Easter, *pedal*). This effect was observed despite variable durations of inter-stimulus intervals, suggesting sensitivity to meter at the word level because meter at the phrase level was irregular. A potential replication of the Magne et al. (2016) study could test whether word-level meter is linked to phonemic segments in perception by measuring a response to an unexpected stress pattern in a list that includes or does not include segmental overlap. For instance, we may expect a greater negativity response in an overlapping list (e.g., candle, candy, canvas, *canteen*) compared to a non-overlapping list (e.g., flower, jacket, llama, *canteen*). If this finding were realized, that would provide evidence in the perceptual domain for the connection of meter and segments.

5.2 CONCLUSIONS

This dissertation investigated the role of metrical spell-out in word lengthening experiments. Metrical overlap did have an effect on word lengthening but only when combined with segmental overlap, indicating that segments may be intimately integrated with stress during phonological encoding. While we did not find support for an abstract metrical structure in words, our results suggest that metrical structure is tied to segmental structure in a critical way. We propose that syllable retrieval or formation is informed by acoustic representations of the target speech sounds. Acoustic similarity of words is greatest with segmental and metrical overlap, and we provide evidence that this degree of similarity leads to increased phonological competition as seen in slower word productions. Our data support the claims that word duration can be primed by recent auditory experience and that phonemic segments and metrical structure are integrated during phonological encoding.

REFERENCES

Aichert, I., Späth, M., & Ziegler, W. (2016). The role of metrical information in apraxia of speech. Perceptual and acoustic analyses of word stress. *Neuropsychologia, 82*, 171–178. https://doi.org/10.1016/j.neuropsychologia.2016.01.009

Arnold, J. E., & Watson, D. G. (2015). Synthesizing meaning and processing approaches to prosody: performance matters. *Language, Cognition and Neuroscience, 30*(1–2), 88–102. https://doi.org/10.1080/01690965.2013.840733

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. https://doi.org/10.1177/00238309040470010201

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

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

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language, 42*(1), 1–22. https://doi.org/10.1006/jmla.1999.2667

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Beirne, M.-B., & Croot, K. (2018). The prosodic domain of phonological encoding: Evidence from speech errors. *Cognition, 177*, 1–7. https://doi.org/10.1016/j.cognition.2018.03.004

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. https://doi.org/10.1016/j.jml.2008.06.003

Berg, T. (1990). The differential sensitivity of consonants and vowels to stress. *Language Sciences, 12*(1), 65–84. https://doi.org/10.1016/0388-0001(90)90024-B

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer (Version 6.0.37). Retrieved from http://www.praat.org/

Boomer, D. S., & Laver, J. D. M. (1968). Slips of the Tongue. *International Journal of Language & Communication Disorders, 3*(1), 2–12. https://doi.org/10.3109/13682826809011435

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Buxó-Lugo, A., Jacobs, C. L., & Watson, D. (in revision). The world is not enough to explain lengthening of phonological competitors. https://doi.org/10.31234/osf.io/6fnwa

Buxó-Lugo, A., Toscano, J. C., & Watson, D. G. (2018). Effects of Participant Engagement on Prosodic Prominence. *Discourse Processes, 55*(3), 305–323. https://doi.org/10.1080/0163853X.2016.1240742

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language, 89*, 68–86. https://doi.org/10.1016/j.jml.2015.12.009

Cholin, J., Levelt, W. J., Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99*(2), 205-35.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology, 6*(2). https://doi.org/10.1515/lp-2015-0008

Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion, 27*(4), 658–684. https://doi.org/10.1080/02699931.2012.732559

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*(2), 292–314. https://doi.org/10.1016/S0749-596X(02)00001-3

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America, 97*(1), 491–504.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037//0033-295X.93.3.283

Fletcher, J. (2010). The Prosody of Speech: Timing and Rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 521–602). Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9781444317251.ch15

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*(5), 489–504. https://doi.org/10.1016/0749-596X(87)90136-7

Fraundorf, S.H., Diaz, M.I., Finley, J.R., Lewis, M.L., Tooley, K.M., Isaacs, A.M., Lam, T.Q., Trude, A.M., Brown-Schmidt, S., & Brehm, L. (2014). CogToolbox for MATLAB [computer software].

Fromkin, V. A. (1971). The Non-Anomalous Nature of Anomalous Utterances. *Language, 47*(1), 27. https://doi.org/10.2307/412187

Fry, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. *The Journal of the Acoustical Society of America, 27*(4), 765–768. https://doi.org/10.1121/1.1908022

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*(1), 35–51. https://doi.org/10.1016/j.jml.2009.09.002

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications, 4*(1). https://doi.org/10.1038/ncomms2430

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Oxford, UK ; Cambridge, Mass., USA: B. Blackwell.

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science, 18*(4), 635–644. https://doi.org/10.1111/desc.12230

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing, 23*(8), 995–1019. https://doi.org/10.1007/s11145-009-9186-6

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics, 31*(3–4), 465–485. https://doi.org/10.1016/j.wocn.2003.09.005

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and Affective Prosody in Children With High-Functioning Autism. *Journal of Speech Language and Hearing Research, 53*(3), 778. https://doi.org/10.1044/1092-4388(2009/08-0127)

Guenther, F.H. (2014). Auditory feedback control is involved at even sub-phonemic levels of speech production. *Language, Cognition and Neuroscience, 29*, 44-45.

Guenther, F.H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics, 72*, 43-53.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23–34.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience, 29*, 2-20.

Holt, C. M., Yuen, I., & Demuth, K. (2017). Discourse strategies and the production of prosody by prelingually deaf adolescent cochlear implant users. *Ear and Hearing, 38*(2), e101–e108. https://doi.org/10.1097/AUD.0000000000000362

Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language, 84*, 37–48. https://doi.org/10.1016/j.jml.2015.05.004

Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology, 61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

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. https://doi.org/10.1037/0278-7393.20.4.824

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*(2), 252–293. https://doi.org/10.1016/0010-0285(92)90009-Q

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language, 67*(3), 311–325. https://doi.org/10.1016/j.jml.2012.07.002

Kahn, J. M., & Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. *Language, Cognition and Neuroscience, 30*(1–2), 103–119. https://doi.org/10.1080/01690965.2013.848989

Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics, 101*, 112–156.

Kenstowicz, M. J. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Kleiner, M., Brainard, D., & Pelli, D., (2007). What's new in Psychtoolbox-3? *Perception 36* ECVP Abstract Supplement.

Kunert, R., & Jongman, S. R. (2017). Entrainment to an auditory signal: Is attention involved? *Journal of Experimental Psychology: General, 146*(1), 77–88. https://doi.org/10.1037/xge0000246

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition, 38*(8), 1137–1146. https://doi.org/10.3758/MC.38.8.1137

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass: M.I.T. Press.

Leong, V., & Goswami, U. (2015). Acoustic-Emergent Phonology in the Amplitude Envelope of Child-Directed Speech. *PLOS ONE, 10*(12), e0144411. https://doi.org/10.1371/journal.pone.0144411

Levelt, W. J. M. (1989). Speaking: from intention to articulation (5. print). Cambridge, Mass.: MIT Press.

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

Lieberman, P. (1960). Some Acoustic Correlates of Word Stress in American English. *The Journal of the Acoustical Society of America, 32*(4), 451–454. https://doi.org/10.1121/1.1908095

Mackay, D. G. (1970). Phoneme Repetition in the Structure of Languages. *Language and Speech, 13*(3), 199–213. https://doi.org/10.1177/002383097001300306

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*(5), 524–545. https://doi.org/10.1016/0749-596X(90)90050-A

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42*(1–3), 181–211. https://doi.org/10.1016/0010-0277(92)90043-H

Motley, M. T. (1973). An analysis of spoonerisms as psycholinguistic phenomena. *Speech Monographs, 40*(1), 66–71. https://doi.org/10.1080/03637757309375781

Myers, B., Lense, M. L., Gordon, R. L. (2019). Pushing the envelope: Developments in neural entrainment to speech suggest a missing link in biological underpinnings of prosody perception. *Brain Sciences, 9*(3), 70.

Myers, B., & Watson, D. (2018). The role of meter in word lengthening. Retrieved from osf.io/zk4qv.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41*(1), 233–243. https://doi.org/10.1016/S0167-6393(02)00106-1

Nooteboom, S. (1973). The tongue slips into patterns. In A. G. Sciarone, A. J. van Essen, & A. A. van Raad (Eds.), *Nomen: Leyden studies in linguistics and phonetics* (pp. 114–132). The Hague: Morton.

O'Seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 57–73. https://doi.org/10.1037//0096-1523.26.1.57

Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language, 78*, 1–17. https://doi.org/10.1016/j.jml.2014.10.003

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences, 27*(2), 169–190; discussion 190-226.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics, 2*, 33-52.

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. Journal of Experimental Psychology. *Human Perception and Performance, 16*(3), 564–573.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 922–939. https://doi.org/10.1037/0278-7393.24.4.922

Rosenbaum, D. A. (1980). Human movement initiation: Specification of arm, direction, and extent. *Journal of Experimental Psychology: General, 109*(4), 444–474.

Rossion, B., & Pourois, G. (2003). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217–236.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143–165. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76–92. https://doi.org/10.1177/0022022101032001009

Schiller, N. O., Jansma, B. M., Peters, J., & Levelt, W. J. M. (2006). Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes, 21*(1–3), 112–140. https://doi.org/10.1080/01690960400001861

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53*(2), 91–127. https://doi.org/10.1016/0010-0277(94)90067-1

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition, 133*(1), 140–155. https://doi.org/10.1016/j.cognition.2014.06.013

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & Cognition, 41*(8), 1200–1211. https://doi.org/10.3758/s13421-013-0332-7

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247.

Shattuck-Hufnagel, Stefanie. (1983). Sublexical Units and Suprasegmental Structure in Speech Production Planning. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 109–136). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4613-8202-7_6

Shattuck-Hufnagel, Stefanie. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In *Motor and* *sensory processes of language*. (pp. 17–51). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Shattuck-Hufnagel, Stefanie. (1992). The role of word structure in segmental serial ordering. *Cognition, 42*(1–3), 213–259. https://doi.org/10.1016/0010-0277(92)90044-I

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

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, 6*, 164–215.

Soderstrom, M. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*(2), 249–267. https://doi.org/10.1016/S0749-596X(03)00024-X

Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington: Indiana University Linguistics Club.

Sulpizio, S., Spinelli, G., & Burani, C. (2015). Stress affects articulatory planning in reading aloud. *Journal of Experimental Psychology. Human Perception and Performance, 41*(2), 453–461. https://doi.org/10.1037/a0038714

Thompson, W. F., & Balkwill, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica, 2006*(158). https://doi.org/10.1515/SEM.2006.017

Tily, H., & Kuperman, V. (2012). Rational phonological lengthening in spoken Dutch. *The Journal of the Acoustical Society of America, 132*(6), 3935–3940. https://doi.org/10.1121/1.4765071

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes, 27*(7–8), 979–1001. https://doi.org/10.1080/01690965.2011.597153

Watson, D. G., Buxó-Lugo, A., & Simmons, D. C. (2015). The Effect of Phonological Encoding on Word Duration: Selection Takes Time. In L. Frazier & E. Gibson (Eds.), *Explicit and Implicit Prosody in Sentence Processing* (Vol. 46, pp. 85–98). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-12961-7_5

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the Time Course of Phonological Encoding. *Journal of Memory and Language, 34*(3), 311–334. https://doi.org/10.1006/jmla.1995.1014

Yiu, L. K., & Watson, D. G. (2015). When overlap leads to competition: Effects of phonological encoding on word duration. *Psychonomic Bulletin & Review, 22*(6), 1701–1708. https://doi.org/10.3758/s13423-015-0828-1

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion, 8*(4), 494–521. https://doi.org/10.1037/1528-3542.8.4.494

APPENDIX A: STIMULUS LIST FOR EXPERIMENT 1

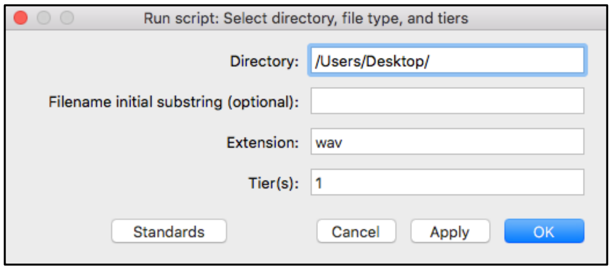
		Prime Word Conditions		
	Target Words	Segmental & Metrical Overlap	Segmental Overlap Alone	Control (No Overlap)
TROCHEES	badger	basket	baguette	sardine
	button	butter	baton	champagne
	candle	candy	canteen	giraffe
	carrot	carriage	carafe	sedan
	castle	casket	cassette	guitar
	locket	locker	lacrosse	baboon
	pickle	picture	pecan	trombone
	racket	rabbit	raccoon	shampoo
	sigma	signal	cigar	grenade
IAMBS	balloon	ballet	ballot	hammer
	bazaar	bassoon	buzzard	coffee
	canoe	canal	cannon	radish
	croissant	crusade	crystal	willow
	delete	deluge	delta	marker
	garage	gazelle	garbage	monkey
	parade	parfait	parrot	shovel
	perfume	percent	person	cauldron
	quartet	corsage	quarter	pillow

Target words were paired with primes that met one of three conditions. Three lists were counterbalanced so that each target appeared only once per list.

APPENDIX B: PRAAT CODING MANUAL

Instructions for Manual Segmentation in Praat

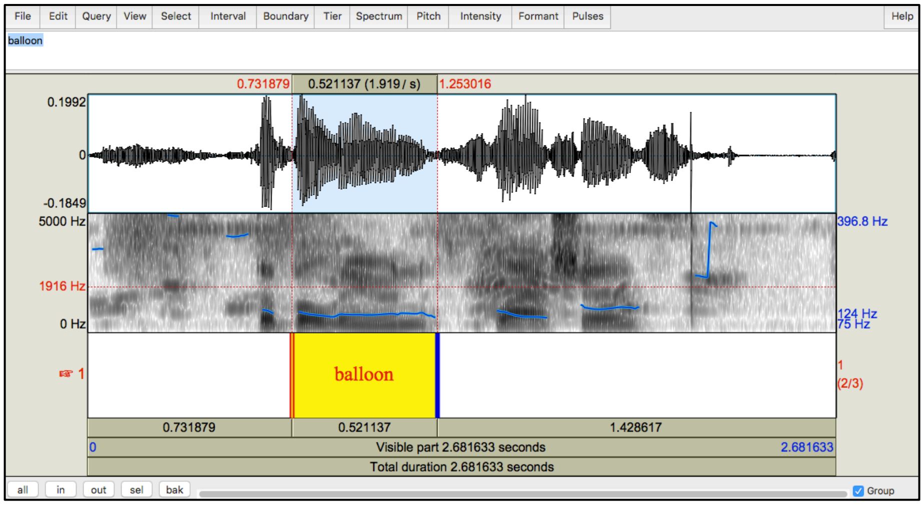
1. Open Praat.
2. In the menu bar, select: Praat > Open Praat script… and open the file called "OpenTextGrid Script".
3. Click "Run" and a pop-up GUI will appear. Enter the folder directory where you will find the sound files for each participant. Click OK.



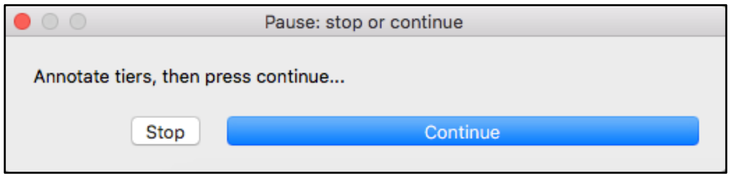
Run script: Select directory, file type, and tiers

Directory: /Users/Desktop/
Filename initial substring (optional):
Extension: wav
Tier(s): 1

4. A window will appear displaying the spectrogram and waveform of the first sound file in the folder. Identify the target word (which is in the file name), and demarcate the start and end points of the word. All target words will be embedded in the phrase: "The *(target)* flashes."
    a) Place a boundary using Command + 1.
    b) Onset: All target words in this experiment begin with a consonant; to identify the start point of the word, look for the change in energy from the end of "The" into the initial consonant. This will often be a sudden burst, but it may also be a gradual shift in formant characteristics. Place a boundary where the consonant begins to form, making sure not to include any portion of the preceding "The". Playback the selection to ensure that you placed the boundary at the onset.
    c) Offset: Identify the end of the word by where the final consonant energy shifts into the /f/ of "flashes". The /f/ will be characterized by high frequency energy with weaker intensity. Place a boundary at the end of the final consonant but before the onset of /f/. Playback the selection to ensure that you placed the boundary at the offset.

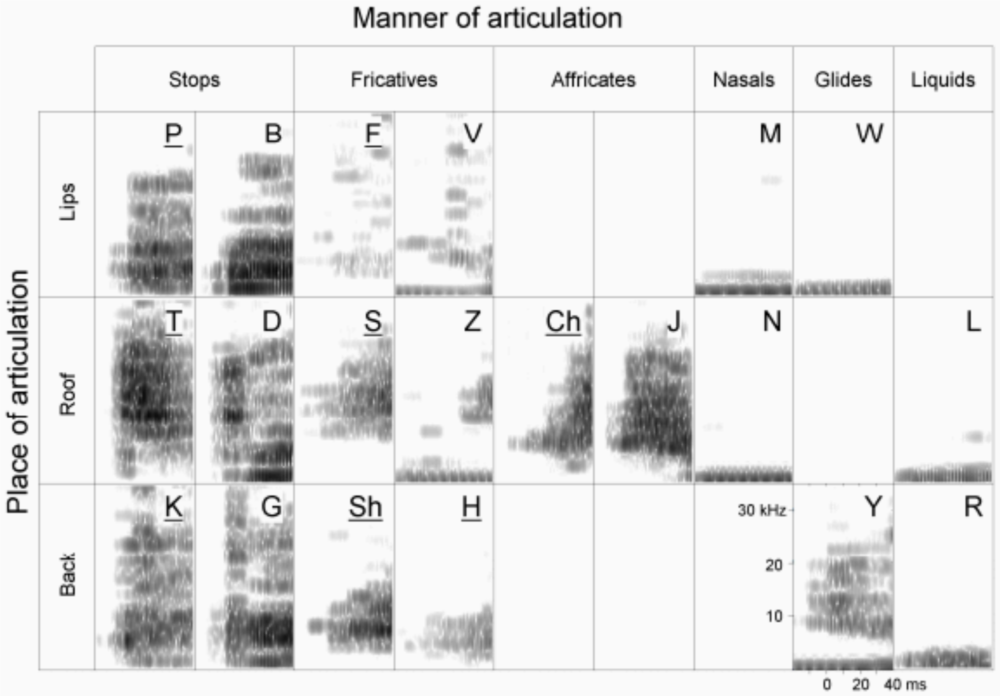


5. Select the word area in the annotation tier, which will highlight yellow. Listen to your selection (by pressing Tab or the play bar directly below the annotation tier), and verify that you have segmented the complete word with nothing extra at either end. Type the target word into the segment as it appears in the file name.
6. When you are satisfied with your segmentation, press Continue in the GUI bar. This will automatically save your TextGrid file to the directory and open the next sound file for you.



Annotate tiers, then press continue...

Here are some example spectrogram representations of consonants to help in your segmentation.



Reproduced from https://www.utdallas.edu/~kilgard/Engineer_NatureNeuroscience_2008_supplement.pdf

APPENDIX C: STIMULUS LIST FOR EXPERIMENT 2

Conditions of Metrical Overlap			
Trochee-Trochee SwSw	**Iamb-Iamb wSwS**	**Trochee-Iamb SwwS**	**Iamb-Trochee wSSw**
arrow error	array arrive	error array	arrive arrow
ballot ballad	ballet balloon	ballad ballet	balloon ballot
beaker beagle	become because	beagle become	because beaker
candle candy	canoe cavort	candy canoe	cavort candle
carrot carriage	career caress	carriage career	caress carrot
college column	collage collide	column collage	collide college
concept concert	concern conceal	concert concern	conceal concept
corner coral	correct corrupt	coral correct	corrupt corner
desert despot	dessert design	despot dessert	design desert
dismal distant	dismay dismiss	distant dismay	dismiss dismal
exit extra	exam excel	extra exam	excel exit
infant info	inform infer	info inform	infer infant
mister missive	mistake misstep	missive mistake	misstep mister
offer office	offend afoot	office offend	afoot offer
parrot parent	parade parole	parent parade	parole parrot
pastor passive	pastel possess	passive pastel	possess pastor
purple purchase	pursuit persist	purchase pursuit	persist purple
recent reason	recite recede	reason recite	recede recent
relic relish	relax relate	relish relax	relate relic
secret seeker	secure secrete	seeker secure	secrete secret

Target phrases consisted of word pairs with initial segmental overlap, and metrical overlap varied across conditions.

APPENDIX D: SPEECH ERRORS CODING MANUAL

Speech Error Analysis Instruction Manual

The following are guidelines for coding speech errors in the tongue twister study. In this study, participants were asked to repeat two-word phrases as fast as possible. Naturally, they made some errors in this task. We will be using these categories to classify the types of speech errors that were made.

Type of Error	**Code**	**Definition**	**Example 1** Target: pastor passive	**Example 2** Target: exit exam
Word Repetition	**WR**	Producing the same word twice in a row	pastor pastor	exam exam
Word Shift	**WS**	Changing the word order in the sequence	passive pastor	exam exit
Phoneme Substitution	**PS**	Misplacing a sound that exists elsewhere in the sequence	pastor pastive	exams exit
Phoneme Insertion	**PI**	Producing a sound not belonging to the sequence	pastor passik	lexam exit
Phoneme Deletion	**PD**	Removing a sound from the sequence	pastor passi	xam exit
Vowel Distortion	**VD**	Producing an incorrect vowel or blend of vowels	pastor possive	exerm exit

We will code these errors in an Excel document named: **SEQ Error Tracking**. The document has a list with all trials that include an error.
The "Errors" column gives a total number of errors in that trial. If there is more than one error in a trial, you will need to insert additional rows for each error. Please copy the trial information into subsequent rows.
In the **"Error Type"** column, you will enter the category code for that error.
In the **"Annotation"** column, you will write what the participant actually said.

Note: In some trials, the participant mispronounces a word or says the wrong word on every repetition (for example: pronouncing *despot* as *despo*). In this case, we cannot classify the utterance as a speech error, but rather we will mark the entire trial as incorrect. In the **"Error Type"** column, place an **X**. In the **"Annotation"** column, write their pronunciation of the target. Do not insert additional rows for each repetition; one row will suffice for this type of trial.

To listen to trials: Open Praat > Open Praat script… > "play Sound" > Run
This will cycle through all .Sound files in the directory. These files are the error clips that were extracted from full-length trials; they include the error, as well as 750 ms on either side of the error.

APPENDIX E: STIMULUS LIST FOR EXPERIMENT 3

		Prime Word Conditions			
	Target Words	Segmental & Metrical Overlap	Segmental Overlap Alone	Metrical Overlap Alone	Control (No Overlap)
TROCHEES	button	butter	baton	sandwich	champagne
	candle	candy	canteen	trumpet	giraffe
	carrot	carriage	carafe	spider	sedan
	castle	casket	cassette	arrow	guitar
	locket	locker	lacrosse	ketchup	baboon
	pickle	picture	pecan	dragon	trombone
	racket	rabbit	raccoon	beaker	shampoo
	sigma	signal	cigar	bucket	grenade
IAMBS	balloon	ballet	ballot	café	hammer
	bazaar	bassoon	buzzard	cocoon	coffee
	canoe	canal	cannon	return	radish
	croissant	crusade	crystal	surprise	willow
	garage	gazelle	garbage	platoon	monkey
	parade	parfait	parrot	cologne	shovel
	perfume	percent	person	roulette	cauldron
	quartet	corsage	quarter	sardine	pillow

Target words were paired with primes that met one of four conditions. Four lists were counterbalanced so that each target appeared only once per list.

APPENDIX F: SUMMARY OF STATISTICAL FINDINGS

Experiment 1

Fixed Effects	***β***	***t***	***p***
*Segmental & Metrical vs Segmental, Control**	0.047	3.729	< 0.001
*Segmental vs Control**	-0.033	-2.446	0.023

Experiment 2

Fixed Effect: Metrical Overlap	***β***	***t***	***p***
*Duration**	-0.039	-2.468	0.016
*Repetitions**	0.087	3.878	< 0.001
*Speech Errors**	-3.13	-5.405	< 0.001
Initial Stress	***β***	***t***	***p***
Duration	0.008	0.510	0.611
Repetitions	0.014	0.514	0.514
Speech Errors	-0.617	-1.103	0.274
Interaction	***β***	***t***	***p***
Duration	0.030	0.933	0.354
Repetitions	-0.067	-1.535	0.129
Speech Errors	1.199	1.039	0.302

Experiment 3

Fixed Effects	***β***	***t***	***p***
*Segmental Overlap**	0.054	5.038	< 0.001
Metrical Overlap	-0.017	-1.714	0.102
*Segmental x Metrical**	-0.038	-2.030	0.062

Statistical findings from three experiments are summarized here. Statistical significance was determined by a *t*-value with an absolute value above 1.96 in a two-tailed test (Baayen, 2008). Significant effects are indicated with an asterisk. In addition, *p*-values are provided for supplementary reference.