

ABSTRACT

Metrical structure (e.g., PREsent vs. preSENT) and segmental structure (e.g., dot vs. pot) are two important cues for differentiating word meanings in English. However, it is not clear whether they are independently or jointly stored in lexical representation. The current study explored this question through the lens of speech planning. In addition, we investigated whether individual differences in working memory capacity (WMC) and music perception skills moderate the effect of word pair characteristics on speech planning difficulty. To study these questions, we adapted the tongue-twister word pair repetition paradigm used by Myers and Watson (2021), and fully crossed metrical overlap, initial stress, and initial segmental overlap. We also calculated the phonological distance between the two words in each word pair. Participants repeated word pairs as quickly as possible while being audio-recorded and completed a series of standardized assessments of WMC and music perception skills. Speech planning difficulty was measured as mean word pair duration, number of correct and complete repetitions, and number of speech errors. We found that speech production was more challenging when the word pair had metrical overlap or segmental overlap, suggesting a joint representation of metrical structure and segmental structure. Speech production was also more challenging when the first syllable was unstressed. There was no moderation effect of individual differences in WMC and music perception skill on the relationship between word pair characteristics and speech planning difficulty. The implications on lexical representation, speech planning, and the connection between music processing and language processing are discussed.

Evidence of Absence? Abstract Metrical Structure in Speech Planning

by

Yang Di

A Thesis

Presented to the Faculty of Mount Holyoke College

in partial fulfillment of the requirements

for the degree of Bachelor of Arts

with Honors

Department of Psychology and Education

Mount Holyoke College

South Hadley, Massachusetts

ACKNOWLEDGEMENTS

First of all, I would like to thank my thesis advisor, Professor Mara Breen, for her enthusiastic support throughout my research journey. There is no way that I would be where I am now without her. Before meeting her, I never knew that prosody is something we can study scientifically. Prosody then led me to psycholinguistics, then meter and rhythm, and now motor control (what?). After knowing her, a question that comes up every so often is: How come she has worked on every single thing that I am interested in? I am excited to continue on this path and further explore the beauty of cognitive science.

I would also like to thank my committee members, Professors Kathy Binder and Tori Day, who knew me since my first semester at Mount Holyoke and continued to guide and witness my growth. My interest in Cognitive Psychology started in Professor Kathy Binder's class, and I may not be a Math major without Professor Tori Day's keen belief that everyone can be a mathematician. Thank you both for always believing in me and cheering me up.

Thank you, Dr. Katerina Drakoulaki, for holding my hand and teaching me so much in the past two years. You have inspired so much academic and personal growth in me. And thank you to everyone in the CAPSlab for being great lab mates and helping me with data collection and cleaning. This project would not have been possible without the many many hours you spent.

I am truly grateful for all the support and encouragement that Dr. Janelle Gagnon had offered. Thank you for your step-by-step guidance and thoughtful feedback on my writing and

presentation. I learned so much from this class, not only about writing and public speaking, but also about giving and receiving peer feedback. I will carry along these skills in my future academic journey. I would also like to thank Janet Crosby for her dedication to supporting everyone in the Department of Psychology and Education.

To my friends in the Honors Thesis Seminar, we made it! This process would have been worth it just to meet and be friends with you. Knowing that I will meet you every Friday is a huge emotional support. The days and nights we spent in the Stimson room became one of my best memories at Mount Holyoke. If I were to describe my thesis writing process, I would say that it smells like Lizzy and Mars's late night delivery; it tastes like the dinner in Applebee's; it feels like the wind that blew away the umbrellas in the balcony of the dining common; and it sounds like six people typing loudly at the same time (I wanted to say that it sounds like Mars opened another energy drink, but I thought I should probably not only list food and drinks).

To my parents and grandparents, I am forever grateful for your love and support. Thank you for encouraging me to pursue the things that excite me and for teaching me perseverance. To my friends, thank you for reminding me how proud of me you are and for tolerating the psycholinguistic jargons I threw at you.

I would like to thank Professor Isabelle Beaudry for showing me the beauty of statistics and getting me started on one of my hobbies - coding in R. All my research projects have benefited from the statistics I learned from you. I also want to thank Mariam Asatryan for the incredible introduction to Linguistics, and Professors Shota Momma, Brian Dillon, and Kristine Yu for being important mentors on my psycholinguistic journey.

Thanks to Mr. and Mrs. Harap for the generous funding that supported this project.

Finally, to science, thank you for all the joys and tears, and excitement and frustrations.

TABLE OF CONTENTS

	Page
Acknowledgements.....	iii
List of Tables	viii
List of Figures.....	x
Introduction.....	1
Lexical Representation.....	2
Individual Differences	8
Individual Differences in WMC	9
Individual Differences in Music Perception	10
Current Study	12
Method	13
Participants.....	13
Materials	13
Procedure	15
Word Pair Repetition Task.....	15
Digit Span Test	16

Operation Span Test.....	16
Rhythmic Contour Test.....	17
Metric Test.....	17
Data Pre-processing	17
Word Pair Repetition	17
Working Memory Capacity	18
Music Perception Skill.....	19
Data Exclusion Criteria.....	19
Analysis.....	20
Hypothesis 1.....	21
Hypothesis 2.....	22
Hypothesis 3.....	22
Results.....	23
Descriptive Statistics.....	23
Word Pair Repetition Measures	23
Individual Difference Measures.....	24
Regression Model Results.....	33
Word Pair Repetition Results.....	33
WMC Results.....	50
Music Perception Skill Results	57

Discussion.....	62
Lexical Representation.....	63
Initial Segmental Overlap Models	63
Phonological Distance Models	65
Comparison Between the Two Sets of Models.....	66
An Alternative Explanation of the Current Results	70
Individual Differences and Speech Planning.....	75
Individual Differences in WMC	75
Individual Differences in Music Perception Skill.....	76
Conclusion	77
Appendix.....	78
References.....	80

LIST OF TABLES

Table	Page
1	The Distribution of Phonological Distance in the Word Pairs 14
2	Mean (SD) for Each Dependent Variable in the Word Pair Repetition Task.....24
3	Mean and Standard Deviation for the Raw Scores in the Individual Difference Tests 25
4	Mean and Standard Deviation for the Raw Scores in the Individual Difference Tests for the 10 Participants Included in the Current Analysis.....29
5	Number of Participants with High Versus Low WMC Scores and Music Perception Scores.....33
6	The Initial Segmental Overlap Model for Mean Word Pair Duration.....34
7	The Phonological Distance Model for Mean Word Pair Duration38
8	The Initial Segmental Overlap Model for Number of Repetitions41
9	The Phonological Distance Model for Number of Repetitions46
10	The Initial Segmental Overlap Model for Number of Errors49
11	The Phonological Distance Model for Number of Errors50
12	The WMC Score Model for Mean Word Pair Duration51
13	The WMC Score Model for Number of Repetitions54
14	The WMC Score Model for Number of Errors56
15	The Music Perception Skill Model for Mean Word Pair Duration58
16	The Music Perception Skill Model for Number of Errors.....60

A1	Word Pairs with Initial Segmental Overlap.....	78
A2	Word Pairs Without Initial Segmental Overlap.....	79

LIST OF FIGURES

Figure		Page
1	Correlation Between Digit Span Test Accuracy Rate and Operation Span Test Accuracy Rate	26
2	Correlation Between Rhythmic Contour Test Accuracy Rate and Metric Test Accuracy Rate	27
3	Correlation Between WMC Score and Music Perception Score	28
4	Correlation Between Digit Span Test Accuracy Rate and Operation Span Test Accuracy Rate for the 10 Participants Included in the Current Analysis	30
5	Correlation Between Rhythmic Contour Test Accuracy Rate and Metric Test Accuracy Rate for the 10 Participants Included in the Current Analysis	31
6	Correlation Between WMC Score and Music Perception Score for the 10 Participants Included in the Current Analysis	32
7	Mean Word Pair Duration by Initial Stress and Initial Segmental Overlap.....	35
8	Mean Word Pair Duration by Initial Segmental Overlap and Metrical Overlap	36
9	Mean Word Pair Duration by Initial Segmental Overlap and Initial Stress.....	37
10	Mean Word Pair Duration by Phonological Distance.....	39
11	Mean Word Pair Duration by Phonological Distance and Initial Stress.....	40
12	Number of Repetitions by Metrical Overlap, by Initial Stress, and by Initial Segmental Overlap	42

13	Number of Repetitions by Initial Segmental Overlap and Metrical Overlap	43
14	Number of Repetitions by Initial Segmental Overlap and Initial Stress	44
15	Number of Repetitions by Initial Segmental Overlap, Metrical Overlap, and Initial stress	45
16	Number of Repetitions by Phonological Distance	47
17	Number of Repetitions by Phonological Distance and Initial Stress	48
18	Mean Word Pair Duration by WMC Score	52
19	Mean Word Pair Duration by Initial Segmental Overlap, Metrical Overlap, and Initial Stress	53
20	Number of Repetitions by WMC Score	55
21	Number of Errors by WMC Score	57
22	Mean Word Pair Duration by Music Perception Score	59
23	Number of Errors by Music Perception Score	61

INTRODUCTION

A written dictionary can include a variety of information about each entry (i.e., word). English dictionaries, for example, list spelling, pronunciation (including phonemes and the stress pattern), meanings, and sometimes example sentences and etymology. The human mental lexicon is like a dictionary in the mind. Similar to dictionaries, we store information for each entry in our mental lexicon. The information stored with each entry is called the lexical representation. This analogy between a written dictionary and a mental lexicon guides cognitive scientists to one of the fundamental questions of language: What information is stored in lexical representations? Is it the same as the information provided in a dictionary? The current study focused on metrical structure (i.e., the stress pattern in words). Specifically, using a tongue-twister word pair repetition paradigm, we aimed to explore whether metrical structure is stored independently in our lexical representation, or, similar to its representation in dictionaries, is stored *with* phonemic information. Additionally, given that different versions of dictionaries may list different information for each word and may have varied organizations of the information, does everyone organize and use their mental lexicon in the same way? The current study explored whether and how individual differences in working memory capacity (WMC) and music perception skills may affect the organization and use of mental lexicon. Understanding these questions would advance our knowledge of how people access and process words in speech production.

Lexical Representation

A phoneme, or segment, is the smallest unit of sound in a language that distinguishes the meaning of one word from another. For example, /p/¹ and /d/ are phonemes in English because *pot* has a different meaning than *dot*. Phonemes are further categorized as vowels and consonants. Syllable is the next large unit of sound that includes minimally a nucleus (most often a vowel), and optional onsets and codas (consonants). For example, while the word *metrical* has 8 phonemes (/m/, /ɛ/, /t/, /ɪ/, /l/, /k/, /ə/, /l/), it only has 3 syllables (/mɛ/, /tɪ/, /kəl/). In addition to phonemes, the metrical structure (i.e., the sequence of stressed and unstressed syllables) is also important for differentiating meanings of words in English (e.g., PREsent vs. preSENT). However, it is unclear whether there is an abstract metrical structure in lexical representation. That is, is metrical structure bound with phonemes, or are metrical structure and segmental structure two independent representations?

One possibility is that there is an abstract metrical structure in lexical representation, such that the planning of metrical structure is independent of the selection of phonemes (Levelt et al., 1999; Roelofs, 1997; Roelofs & Meyer, 1998). In other words, metrical structure and segmental structure are two independent full representations. According to this account, metrical structure and individual phonemes are retrieved in parallel. The retrieved metrical structures specify whether each syllable is stressed or not, and the retrieved phonemes are distributed to these syllables. That is, the planning of word sounds is finished in two steps: The first step is the parallel retrieval of metrical structure and phonemes, and the second step is the association of the metrical structure and phonemes, as described in the word encoding via activation and

¹ Following standard notation, phonemes as defined by the International Phonetic Alphabet (IPA) are indicated using slash notation.

verification (WEAVER) model (Roelofs, 1997). For example, when we access the word “arrow,” we know there are two syllables, with the first syllable being stressed and the second syllable being unstressed. We also know that the segments, or phonemes, of this word are /ɛ/, /r/, and /oʊ/. Then we slot these phonemes into the syllable positions, forming /ɛ/ and /rou/. Finally, the first syllable is stressed and the second syllable is unstressed, resulting in the final form of /'ɛrou/, where ' marks the location of the stress. Support for this account came from the literature on the Tip-of-the-Tongue (TOT) phenomenon during which people fail to retrieve the phonological information of a word (i.e., cannot produce the word). Brown and McNeil (1966) found that during the TOT state, participants were generally accurate about the number of syllables in the word and the location of primary stress. Therefore, while the segmental structure is not available, the metrical structure may still be activated, suggesting that segmental structure and metrical structure may be retrieved independently.

To establish abstract representations in language processing, previous literature has used priming effects. Priming effects refer to the phenomenon that people tend to use a representation that is recently encountered, and that the processing of an upcoming item (i.e., the target) would be facilitated if it shares a full representation with a recently encountered item (i.e., the prime). For example, Bock (1986) argued that there is an abstract syntactic representation by showing a syntactic priming effect in speech production. She found that when people were asked to use one sentence to describe an event depicted in a picture, the probability of using a particular syntactic construction increased when the prime sentence used the same syntactic construction. Crucially, priming should only occur when the prime and the target have a full overlap in a representation, but not when they have a partial overlap in a representation. For instance, Tooley and colleagues concluded that neither intonational boundary nor pitch accent is a full representation on its own

because they were not primable (Tooley et al., 2014, 2018). Therefore, if there is an abstract metrical structure and an abstract segmental structure in lexical representation, then we should observe priming effects of metrical overlap and segmental overlap, respectively, in language processing.

An alternative possibility is that there is no abstract metrical structure in lexical representation. That is, metrical structure is bound to segmental structure, meaning that metrical structure and segmental structure are only partial representations on their own. Therefore, the stressed version and the unstressed version of a syllable would be stored as two independent units, and the stressed or unstressed version of two different syllables would also be stored as independent units. In other words, when we access the word “arrow”, we know that the first syllable is the stressed version of /ε/ and the second syllable is the unstressed version of /rou/. Concatenating them results in the final form of /'εrou/. If metrical structure and segmental structure are indeed bound together, then there should be no priming effect when there is only metrical overlap or only segmental overlap. However, when two words share both metrical structure and segmental structure, even only for a part of the words, the activation of one word would activate the other one, leading to competition between the two words and thus slowing down the word processing. In other words, there should be an interference effect of the co-occurrence of metrical and segmental overlaps.

These hypotheses can be tested through the lens of speech planning, during which messages are translated into motor commands that produce sounds. A simplified version of the WEAVER++ model (an expanded version of the WEAVER model; Levelt et al., 1999) roughly includes four stages in speech planning: conceptualization, word selection, phonological encoding, and articulation. Conceptualization is the process of forming a message; word

selection is the process of selecting abstract word representations; phonological encoding is the process of retrieving phonological representation from the mental lexicon; and articulation is the process of moving muscles to make sounds. Similar to how reaction time can be a good indicator of processing difficulty (Schneider et al., 2011), word duration and speech onset latency have been used as indices of speech planning difficulty (Meyer & van der Meulen, 2000; Watson et al., 2015). Hence, if metrical structure is planned independent of segments (phonemes), then the word duration should be shorter when two words share metrical structure or segmental structure. On the contrary, if metrical structure is bound to the representation of segments, then the word duration should be longer when two words share metrical structure and segmental structure.

Previous studies exploring the effect of metrical congruence on speech planning difficulty have shown mixed results. On one hand, previous studies have observed priming effects of metrical overlap and segmental overlap, suggesting that metrical structure and segmental structure are independently planned. Priming effects of metrical regularity have been found in a non-word sequence repetition experiment, such that there were faster productions, fewer hesitations, and fewer sequence errors in the regular metric condition than in the irregular metric condition (Tilsen, 2011). Priming effects of segmental overlap have been found in picture-naming experiments (Damian & Dumay, 2007). When participants were presented with a picture and a distractor word, and were asked to describe the picture using a determiner-adjective-noun phrase (e.g., the blue bell), the speech onset latency was shortened when the distractor was phonologically related to the noun (i.e., the distractor and the noun shared a minimum of the initial consonant cluster and sometimes also the following vowel). When a distractor word was not present, the onset latency for the noun word was also shortened when the adjective and the noun shared the initial phoneme. These priming effects of metrical overlap and

segmental overlap supported the existence of an abstract representation of metrical structure and an abstract representation of segmental structure respectively.

On the other hand, interference effects of segmental overlap and metrical overlap have also been found in previous event-describing experiments, suggesting that segmental structure and metrical structure are jointly planned. In Yiu and Watson (2015), participants were asked to describe events where the first object (the prime word) shrunk and the second object (the target word) flashed. The word duration of the target word was longer when the target word phonologically overlapped with the prime word, suggesting an interference effect of segmental overlap. In a similar experiment with a 2 (initial segmental: overlap vs. no overlap) x 2 (metrical structure: overlap vs. no overlap) design, Myers and Watson (2019) showed that there was greater word lengthening when there were both initial segmental and metrical overlaps than when there was only initial segmental overlap, and that there was no word lengthening effect when there was only metrical overlap. The observed interference effect of concurrent initial segmental and metrical overlaps suggest that metrical structure and phoneme are closely bound. Moreover, because there was word lengthening when there was only initial segmental overlap but not when there was only metrical overlap, their results suggest that phonemes may play a more important role in speech planning than metrical structure does.

Given these mixed results, further research is needed to clarify the effects of metrical overlap and segmental overlap on speech planning. Therefore, the current study further explored the planning of metrical structure and segmental structure, specifically during the phonological encoding stage, aiming to provide insights into not only the planning of metrical structure and segmental structure during speech production but also their encoding in the mental lexicon.

To focus on the effects of metrical overlap and segmental overlap on phonological encoding, we need to control conceptualization, word selection, and articulation. Removing conceptualization and word selection from the speech planning process can be achieved by asking people to read aloud written text, instead of producing spontaneous speech. A tongue twister paradigm can further elicit pure pre-articulatory errors (Wilshire, 1999), suggesting that articulation difficulty can be controlled in this paradigm. Myers and Watson (2021) thus used tongue-twister word pairs to explore the effect of metrical overlap on speech planning. In the experiment, participants saw a word pair for 4 seconds, then the word pair disappeared, and participants repeated the word pair as many times as possible for 8 seconds. In addition to metrical overlap, the authors also manipulated the initial stress (i.e., whether the first syllable was stressed or unstressed). Aichert and colleagues (2016) found that German native speakers produced iambic meters (weak-strong) with longer durations and more errors than trochaic meters (strong-weak). One explanation for this effect is that there are more trochaic words than iambic words in German (Féry, 1998), and thus German native speakers are better at producing trochaic words. Similarly, in English, there are more trochaic words than iambic words and trochaic words are used more frequently in daily speech (Cutler & Carter, 1987). Therefore, it is important to control for the effect of initial stress on speech production. By manipulating metrical overlap and initial stress, Myers and Watson (2021) found that word pairs with metrical overlap were produced fewer times, and with longer duration and more errors, suggesting that there is no abstract metrical structure in lexical representation and that metrical structure is bound to phonemes in speech planning.

However, there are two limitations of this study. First, while Myers and Watson (2021) showed an interference effect when there is both metrical and initial segmental overlap, the

effect of initial segmental overlap alone remained untested because all word pairs they used shared the first segment. Second, they only controlled the overlap of the first segment, ignoring other segments in the words. Although the cohort model proposed that only the word-initial phonemes play a crucial role in word recognition (Marslen-Wilson & Welsh, 1978), other models suggest that all segmental overlap, regardless of position, affects word activation (e.g., the TRACE model, McClelland & Elman, 1986). For example, words that do not share the word-initial phoneme but rhyme with the heard word are also activated (Allopenna et al., 1998). Thus, aiming to replicate and extend the results from Myers and Watson (2021), the current study (1) fully crossed metrical overlap, initial stress, and initial segmental overlap, and (2) calculated the phonological distance between the two words in each word pair to capture mismatches of phonemes in all positions in the words, and included this phonological distance as a predictor of the difficulty of speech planning. This way, we were able to examine how metrical structure and segmental structure are stored in lexical representation, and assess the effects of initial segmental overlap and phonological distance on word production separately.

Individual Differences

In addition to the open question of how metrical structure and segmental structure are stored in lexical representations, it is also unclear whether all people use their mental lexicon in the same way. In other words, do people access their lexical representations through the same informational pathways? To explore this question, we tested whether and how individual differences in working memory capacity (WMC) and music perception skills may affect individuals' use of lexical representation in speech planning.

Individual Differences in WMC

WMC has been shown to affect various aspects of language processing. Just and Carpenter (1992) observed that people with better WMC are better at integrating pragmatic (i.e., contextual meaning) and syntactic (i.e., sentence structure) information, and, of particular interest to the current study, better WMC allows people to maintain multiple interpretations. During our tongue-twister word pair repetition task, if people cannot simultaneously hold the production plans for both words, then they may need to plan a word and then remap the speech plan to produce a similar-sounding word for every alternation of words (Rosenbaum et al., 1986), making the speech production more challenging. However, if people with better WMC can hold multiple speech plans at the same time, then they would not need to remap the speech plans as much as people with worse WMC do, and thus speech production would be easier.

Another possible way for WMC to affect speech planning is that people with better WMC may have a larger processing span. Swets et al. (2007) showed that people with lower working memory span tended to insert breaks in large segments of texts due to the limited working memory. Similarly, in speech production experiments, participants with better WMC produced longer prosodic phrases, suggesting that they had a larger planning span (Bishop, 2021; Bishop & Intlekofer, 2020). If having a better WMC indeed allows a larger speech planning scope, then participants with better WMC might even be able to have a single speech plan for the whole word pair, completely eliminating the need for remapping.

Therefore, we hypothesized that people with better WMC would be better able to maintain both production plans and have less need for remapping, resulting in less difficulty producing challenging word pairs. That is, speech planning difficulty will be less affected by metrical overlap and phonological overlap for people with better WMC.

Given the previous work showing that simple span and complex span measures of WMC account for different variance in language processing (Daneman & Carpenter, 1980), we aggregated results from a digit span test (a simple span) and an operation span test (a complex span) to explore whether the effects of metrical overlap and phonological overlap on speech planning change as a function of WMC.

Individual Differences in Music Perception

Music and language show great similarities (Heffner & Slevc, 2015). For example, music and language can both be hierarchically described, with meter being an important aspect in both musical and prosodic hierarchies. In language, phonemes are grouped to form a syllable, one stressed syllable and optional unstressed syllables form a metrical foot, metrical feet form a word, words form a phrase, and phrases form a sentence. Similarly, in music, notes are associated with beats, beats form measures, and measures form phrases. Music and language also share similar phenomena. For instance, in both music and speech, elements at the phrase-final position tend to be lengthened (e.g., Klatt, 1975; Large & Palmer, 2002).

These parallels between music and language have led scientists to investigate the cross-domain interaction between the processing of music and language. Besson and Schön (2001) discussed the convergent evidence suggesting that similar cognitive mechanisms were involved in harmonic processing in music and syntactic processing in language, and that temporal structural violation has similar effects on language and music processing. In addition, a lower accuracy or a longer reaction time was usually observed during the simultaneous processing of language syntax and music syntax (Fedorenko et al., 2009; Slevc et al., 2009), suggesting that similar cognitive mechanisms were involved in the processing of language syntax and music syntax. Sun and colleagues (2024) found that musical melodies can prime implicit prosodic

contours (i.e., imagined speech melodies), leading to different attachment judgments for globally ambiguous phrases (e.g., *only cars and trucks with permits are allowed*). Moreover, previous work showed that listening to metrically unexpected sung sentences slowed down lexical processing: A longer reaction time was needed to decide if a visually presented string of letters was a real word or nonword after listening to sung sentences where the stressed syllables of the lyrics mismatched the strong metrical position in music, compared to after listening to sung sentences where the stressed syllables of the lyrics matched the strong metrical position in music (Gordon et al., 2011).

Motivated by these robust connections between language and music processing, previous work has explored the relationship between an individual's music skill and language skill. Slevc and Miyake (2006) found that better music ability, as measured by chord analysis, pitch change, tonal memory, and tonal memory production, predicted second language phonological skill even after controlling for other important factors (e.g., motivation for second language use, second language exposure, phonological short term memory, etc.). Additionally, compared to non-musicians, musicians were more sensitive to metrical incongruence in the final words of sentences, suggesting that the ability to process metrical structure in music is correlated with the ability to process metrical structure in language (Marie et al., 2011).

Hence, we hypothesized that people with better music perception skills would be more affected by metrical overlap in speech planning. To test this hypothesis, we used the rhythmic contour task and the metric task of the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz et al., 2003; Vuvan et al., 2018) to measure music perception skill, and assessed whether it moderates the effect of metrical overlap on speech planning.

Current Study

The purpose of the current study was to investigate how metrical structure and phonemic structure are stored in lexical representations, and whether individual differences in WMC and music perception skill moderate the relationship between word pair characteristics and speech planning difficulty. We hypothesized that (1) if there are two independent representations for metrical structure and phonemes, then speech production should be easier when the two words share metrical structure or segmental structure, as evidenced by shorter word pair duration, more complete repetitions, and fewer speech errors; if they are bound in lexical representation, then speech production should be more challenging when two words share metrical structure and segmental structure, as evidenced by longer word pair duration, fewer complete repetitions, and more speech errors; (2) Metrical overlap and phonological overlap would have a smaller effect on speech planning for people with better WMC, as evidenced by an interaction effect between WMC and metrical overlap, and an interaction effect between WMC and initial segmental overlap or phonological distance; (3) Metrical overlap would have a stronger effect on speech planning for people with better music perception skill, as evidenced by an interaction effect between music perception skill and metrical overlap.

To test these hypotheses, we adapted the tongue-twister word pair repetition paradigm from Myers and Watson (2021) and fully crossed metrical overlap, initial stress, and initial segmental overlap. Moreover, we calculated the phonological distance between the two words in each word pair. We used a simple digit span task and an automated operation span task (Conway et al., 2005; Unsworth et al., 2005) to measure individual differences in WMC. We used the rhythmic contour task and the metric task from the MBEA (Peretz et al., 2003; Vuvan et al., 2018) to measure individual differences in music perception skill.

METHOD

Participants

Ten participants (age range: 19-23 years old, $M = 20.4$) contributed to the current analysis. We recruited 61 adult participants (age range: 18-33 years old, $M = 19.83$) who are native speakers of American English, meaning that they had all been speaking English in the United States since at least the age of 5. They were all Mount Holyoke College students, recruited through SONA and word of mouth. They were given the choice of being compensated with two SONA credits or \$10 for their participation. Among the 61 participants, three were excluded from the final analyses: two were excluded because they did not complete the full experiment and one was an outlier for the digit span test (more than three standard deviations away from the mean digit span raw score).

Materials

The word pair repetition task was implemented on FindingFive (FindingFive Team, 2023), an online experiment platform. In the word pair repetition task, we fully crossed metrical overlap (overlap vs. non-overlap), initial stress (strong vs. weak), and initial segmental overlap (overlap vs. non-overlap), resulting in eight conditions. There were 20 word pairs in each condition (i.e., 160 experimental items in total; see Appendix for a list of all experimental items) and three additional word pairs as practice items, summing up to 163 items in total. Each item was coded in an individual trial. The 80 word pairs in the Initial Segmental Overlap conditions were adopted from Myers and Watson (2021). The 80 word pairs in the No Initial Segmental

Overlap conditions were constructed by rearranging the word pairs in the Initial Segmental Overlap conditions so that none of the word pairs shared the first segment. Position for each word was counterbalanced: Each word appeared twice in the first word position and twice in the second word position. The position was not counterbalanced for each word pair (e.g., participants only read *beaker beagle* but not *beagle beaker*) to keep the experiment duration manageable for participants.

To calculate the phonological distance between the two words for each word pair, we used the Levenshtein distance (Levenshtein, 1966), which indexes how different two strings are: the larger the value, the larger the difference. Levenshtein distance was calculated as the number of edits needed to make two strings the same. For example, the Levenshtein distance between “cat” and “hat” is one because only one edit (change “c” to “h”) is needed to make the two strings the same. We calculated Levenshtein distance for the International Phonetic Alphabet (IPA) transcriptions of the two words in each word pair, quantifying the distance between the sounds of the two words. To avoid high collinearity between metrical overlap and phonological distance, we did not include stress markers in the IPA transcriptions. The distribution of phonological distance in the word pairs is shown in Table 1.

Table 1

The Distribution of Phonological Distance in the Word Pairs

Phonological Distance	1	2	3	4	5	6	7
Number of word pairs	4	20	33	30	45	26	2

The working memory task included a digit span test and an automated operation span test, both coded in E-prime 2.0.10. The digit span test included all digits from one to nine. The

operation span test was downloaded from <https://englelab.gatech.edu/standardtasks.html> and the letter pool includes F, P, Q, J, H, K, T, S, N, R, Y, L (Conway et al., 2005; Unsworth et al., 2005).

The music perception test included the rhythmic contour test and the metric test in MBEA (Peretz et al., 2003; Vuvar et al., 2018). Both the rhythmic contour test and the metric test were implemented on FindingFive (FindingFive Team, 2023).

Procedure

All participants completed the experiment individually, under the guidance of an experimenter who had completed ethics training and knew the purpose of the study. Before starting the experiment, participants signed the informed consent form to acknowledge their voluntary participation and their consent for audio recording during the word pair repetition task. Participants then entered a quiet room and completed all the tasks there. They first finished the word pair repetition task, followed by the digit span test, the operation span test, the rhythmic contour test, the metric test, and filled out a questionnaire to provide demographics information and music background information at the end. The experiment took approximately 2 hours to complete.

Word Pair Repetition Task

During the word pair repetition task, participants wore a head-mounted microphone. For each trial, participants first saw the word pair for 4 seconds. Then the word pair disappeared and the word “GO!” showed up on the screen. When they saw the word “GO!” participants started repeating the word pair as quickly as possible for 8 seconds until the word “STOP” appeared on the screen. They were encouraged to take breaks if needed, and when they were ready, they then pressed “continue” to proceed to the next trial. There were 163 trials in total (three practice trials

and 160 experimental trials). All experimental trials were presented in a fully randomized order for each participant.

Digit Span Test

During the digit span test, participants listened to sequences of digits. They needed to type the digit in the correct order. There were six blocks, and each block contained five sequences of the same length. The number of digits per sequence ranged from three to eight. Every participant started with three-digit sequences. If participants got at least two sequences correct for the block containing the sequences of x digits, they would advance to the block with sequences of $x + 1$ digits. Otherwise, they would start the block with sequences of $x - 1$ digits, until they got to the block with one-digit sequences. For example, if a participant correctly recalled all digits in the correct order for three sequences of three-digits, they would advance to the block of four-digit sequences; if a participant only correctly recalled one 3-digit sequence, they would proceed to the block of two-digit sequences. The highest possible raw score was 165 (i.e., $(3 + 4 + 5 + 6 + 7 + 8) * 5$).

Operation Span Test

For each trial of the operation span test, the first screen had a math problem (e.g., “ $(9 / 3) - 2$ ”), the second screen had a proposed answer to the problem and asked the participant whether the answer was correct or not (e.g., “10? True or False”), and the third screen had a letter (e.g. “K”). After a random number of trials, participants were asked to recall all the letters in the order they appeared. The number of trials per block (i.e., the letter sequence length) ranged from three to seven. There were three blocks for each letter sequence length, summing up to 15 blocks in total and 75 letters in total.

Rhythmic Contour Test

During the rhythmic contour test, participants listened to a series of melody pairs and they decided whether the two melodies in each pair were the same or different. There were two example trials, during which participants received feedback and could listen to the melody pairs as many times as needed. There were 31 experimental trials, during which participants did not receive feedback and could only listen to each melody pair once.

Metric Test

During the metric test, participants listened to a series of sequences of melodies, presented one at a time, and decided if each melody was a waltz or a march. They were told that a waltz rhythm is in three with a strong beat followed by two weak beats (e.g., **1**, 2, 3, **1**, 2, 3...) whereas a march rhythm is in two with a strong beat followed by a weak beat (e.g., **1**, 2, **1**, 2...). There were four example trials, during which participants received instructions from the experimenter about how to differentiate between a waltz and a march, and received feedback on their answer. Each example audio could be played as many times as the participant needed. There were 30 experimental trials, during which participants did not receive any instruction nor feedback, and they could only listen to the audios once.

Data Pre-processing

Word Pair Repetition

The word pair repetition data were annotated using Praat (Boersma & Weenink, 2023). Ten coders independently coded a subset of the word pair repetition data. For every trial, a textgrid was created to mark each correct and complete word pair repetition and each error. We then counted the number of correct and complete word pair repetitions, and the number of errors, and transcribed the correct production. For example, if the word pair is “arrow error” and the

participant said “arrow error arrow error arrow error error arrow error,” then the number of correct and complete repetitions would be four because there were four correct and complete repetitions of the word pair “arrow error.” The number of errors would be one because there was an additional “error” that was not paired with an “arrow.” The transcription would be “arrow error arrow error arrow error arrow error.” The transcription was then used to create a text file. We also cut the errors and incomplete word pair repetitions from the original sound file. The text file and the cleaned sound file were then paired to run forced alignment using Montreal Forced Aligner (McAuliffe et al., 2017), which provided us with time intervals for each word. We then extracted the duration for each word pair repetition in Praat (Boersma & Weenink, 2023) and calculated the mean word pair duration for each word pair for each participant.

Working Memory Capacity

We composed the digit span raw score and the operation span score to capture individual differences in working memory capacity.

Digit Span. The digit span raw score was calculated as the total number of digits correctly recalled in the correct position. The final digit span score was calculated as the digit span raw score divided by the total number of digits presented to the participant, as in (1).

$$\text{Digit Span Score} = \frac{\text{total number of digits correctly recalled in the correct position}}{\text{total number of digits presented}} \quad (1)$$

Operation Span. The operation span raw score was calculated as the total number of correct letters recalled in the correct position (i.e., the partial-credit load score) as suggested by Conway et al. (2005) and used in Moss et al. (2024). We calculated the scores using the `raw_ospans()` and `score_ospans()` functions in the `englelab` R package (Tsukahara, 2022). The final operation span score was calculated as the operation span raw score divided by the total number of letters presented (i.e., 75), as done in Conway et al. (2005). See (2) for the equation.

$$\text{Operation Span Score} = \frac{\text{total number of letters correctly recalled in the correct position}}{75} \quad (2)$$

WMC Score. The composite score for the two working memory tests was calculated as the average of the digit span score and the operation span score, as in (3).

$$\text{WMC Score} = \frac{\text{Digit Span Score} + \text{Operation Span Score}}{2} \quad (3)$$

Music Perception Skill

The rhythmic contour test score and the metric test score were separately calculated as the number of correct responses to the experimental trials. We then calculated an accuracy rate for each of the two tests. The composite score for the two music perception tests was calculated as the average of the accuracy rates for the two tests, as in (4).

$$\text{Music Perception Score} = \frac{\text{Rhythmic Contour Accuracy Rate} + \text{Metric Accuracy Rate}}{2} \quad (4)$$

Data Exclusion Criteria

A word pair repetition trial would be excluded from analysis if the participant consistently mispronounced either word incorrectly throughout the whole trial. Based on this criterion, 46 trials were excluded. A single word pair repetition would be excluded from analysis if the repetition duration was at least 2.5 standard deviations away from the mean repetition duration for this specific trial for this participant. We excluded 15 repetitions based on this criterion. Given our way of detecting outliers, we also excluded all trials with only one correct and complete repetition because we were not able to calculate the standard deviation of the word pair durations for this trial and thus could not detect outliers. Six additional trials were excluded based on this criterion. Therefore, a total of 52 out of 1600 trials (3.25%) were excluded and, within the remaining 1548 trials, 15 out of 11198 repetitions (0.13%) were excluded.

Analysis

The current study aimed to examine whether word pair characteristics and their interactions predict the difficulty of speech planning, and whether individual differences in WMC and music perception skill interact with the word pair characteristics to predict the difficulty of speech planning. Specifically, the word pair characteristics included metrical overlap, initial stress, initial segmental overlap, and phonological distance; the WMC score was the average of the digit span score and the operation span score; and the music perception score was the average of the accuracy rates of the rhythmic contour test and metric test. The difficulty of speech planning was measured as mean word pair duration, number of correct and complete repetitions, and number of speech errors. We ran multiple linear regression analyses and t-tests using the stats package (R Core Team, 2023) in R.

Metrical overlap, segmental overlap, and initial stress were all binary variables. Phonological distance was a ratio variable that ranged from one to seven. We median splitted the WMC score and the music perception score to make them binary. For example, a participant whose WMC score was among the highest five scores was coded as having a high WMC score, and a participant whose WMC score was among the lowest five scores were coded as having a low WMC score. All variables were centered by subtracting the mean from each value.

By definition, initial segmental overlap and phonological distance account for similar aspects of the word pair characteristics: How similar the segments in the two words were. Thus, they would be highly correlated: Word pairs with initial segmental overlap would have smaller phonological distance than word pairs with no initial segment overlap. Therefore, to avoid multicollinearity in the regression models, we fitted two groups of models separately: the initial segmental overlap models and the phonological distance models.

Hypothesis 1

Hypothesis 1 stated that if metrical structure and segmental structure are independently planned, then speech production should be easier when there is initial segmental overlap or metrical overlap, as evidenced by shorter mean word pair duration, more repetitions, and fewer errors; if metrical structure and segmental structure are jointly planned, then speech production should be more challenging when there is initial segmental overlap and metrical overlap, as evidenced by longer mean word pair duration, fewer repetitions, and more errors.

To test this hypothesis, the initial segmental overlap models included metrical overlap, initial stress, initial segmental overlap, and all possible interactions as predictors for each dependent variable: mean word pair duration, number of complete repetitions, and number of speech errors. The phonological distance models included metrical overlap, initial stress, phonological distance, and all possible interactions as predictors for each dependent variable: mean word pair duration, number of complete repetitions, and number of speech errors. We then assessed which predictors were significant in each model.

We also compared the initial segmental overlap models to the phonological distance models using their adjusted R^2 values. The model with a higher adjusted R^2 value would be preferred because the model explained more variance in the dependent variable.

For both mean word pair duration and number of repetitions, the initial segmental overlap model had a higher adjusted R^2 than the phonological distance model. In addition, because phonological distance was not a controlled variable, the sample size across different phonological distance values varied. Therefore, we used the initial segmental overlap models as the basic models when we tested the hypotheses about individual differences. That is, the

respective main effects and interaction effects were only added to the initial segmental overlap models.

Hypothesis 2

Hypothesis 2 stated that people with better WMC should be less affected by metrical overlap and phonological overlap. To test this hypothesis, we added WMC score, the interaction between WMC score and metrical overlap, the interaction between WMC score and initial segmental overlap, and the three-way interaction between WMC score, metrical overlap, and initial segmental overlap as predictors in the initial segmental overlap model. We then assessed whether these added variables were significant predictors for each dependent variable: mean word pair duration, number of complete repetitions, and number of speech errors.

Hypothesis 3

Hypothesis 3 stated that people with better music perception skill would be more affected by metrical overlap. To test this hypothesis, we added music perception score and the interaction between music perception score and metrical overlap as predictors in the initial segmental overlap model. We then assessed whether these added variables were significant predictors for each dependent variable: mean word pair duration, number of complete repetitions, and number of speech errors.

RESULTS

Descriptive Statistics

Word Pair Repetition Measures

Mean and standard deviation of mean word pair duration, number of correct and complete word pair repetitions, and number of speech errors are reported in Table 2. Across all conditions, the mean word pair duration was about 866 ms, the number of repetitions was about 7.23, and the number of errors was about 0.68. In general, participants produced very few errors. Only 603 out of 1548 trials included in the analysis had at least one error. For the 603 trials with errors, the mean number of errors was only 1.75.

Table 2*Mean (SD) for Each Dependent Variable in the Word Pair Repetition Task*

Segmental overlap	Metrical overlap	Initial stress	Mean word pair duration (ms)	Number of repetitions	Number of errors
Overlap	Overlap	S	901.90 (8.87)	6.72 (0.10)	0.80 (0.08)
		w	884.77 (10.37)	6.93 (0.11)	0.71 (0.08)
	No overlap	S	867.64 (8.57)	7.38 (0.12)	0.62 (0.09)
		w	876.70 (8.86)	7.19 (0.11)	0.66 (0.09)
No overlap	Overlap	S	824.15 (8.13)	7.70 (0.11)	0.70 (0.08)
		w	872.98 (8.45)	7.11 (0.09)	0.71 (0.07)
	No overlap	S	838.69 (8.62)	7.56 (0.11)	0.59 (0.07)
		w	865.67 (8.15)	7.32 (0.11)	0.67 (0.08)

Individual Difference Measures

Individual Difference Results for All 58 participants. The descriptive statistics for the individual difference measures for all 58 participants are presented in Table 3.

Table 3

Mean and Standard Deviation for the Raw Scores in the Individual Difference Tests

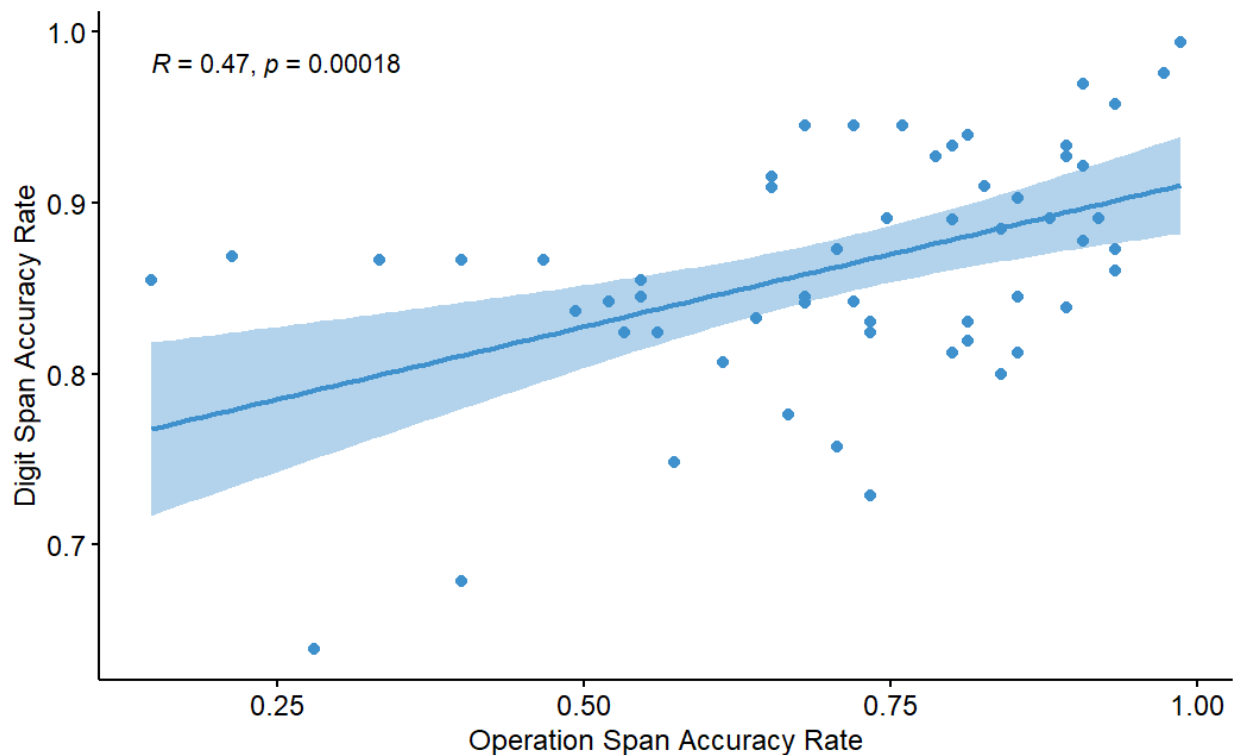
Test	Mean	Standard deviation
Digit Span Test (165)	140	13.4
Operation Span Test (75)	53.3	14.5
Rhythmic Contour Test (31)	25.4	3.44
Metric Test (30)	26.1	4.75

Note. The number in parentheses after the test name indicates the highest score a participant might get on this test.

There was a positive correlation between participants' performances on the digit span test and the operation span test ($r = 0.47$, $p < 0.001$; Figure 1), suggesting that the tests were able to capture individuals' working memory capacity. In addition, the correlation was not close to 1, indicating that each of the two working memory tests captured different aspects of working memory. Thus, the composite WMC score would be a better measure of individual WMC than either one of the test scores.

Figure 1

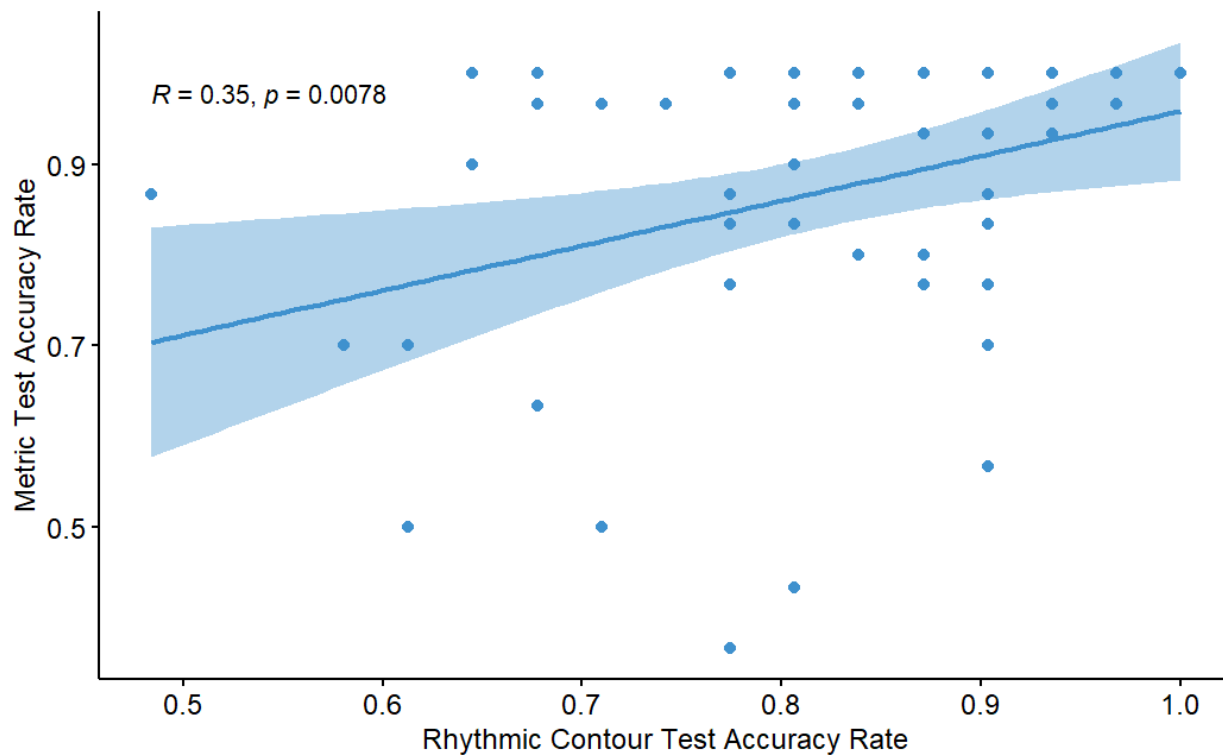
Correlation Between Digit Span Test Accuracy Rate and Operation Span Test Accuracy Rate



There was a positive correlation between participants' performance in the rhythmic contour test and the metric test ($r = 0.35, p = 0.008$; Figure 2), suggesting that the tests were able to capture individuals' music perception skill. Moreover, despite being significant, the correlation was not strong, indicating that the two test scores accounted for different aspects of music perception skill. In addition, many participants were at ceiling on the metric test, suggesting that the metric test alone is not good at uncovering individual differences. Thus, the composite music perception score would be a better measure of individual music perception skill than the accuracy rate for either of the tests.

Figure 2

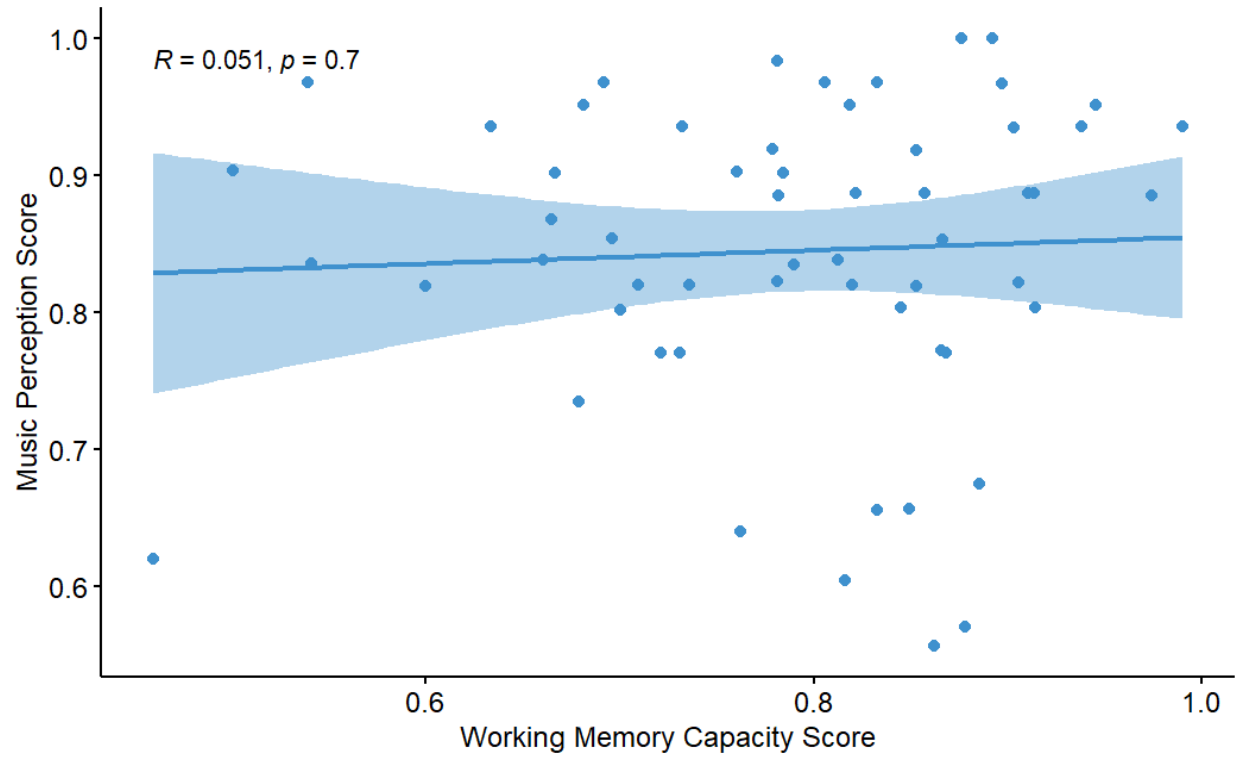
Correlation Between Rhythmic Contour Test Accuracy Rate and Metric Test Accuracy Rate



There was no correlation between an individual's WMC score and music perception score ($r = 0.051, p = 0.7$; Figure 3), suggesting that these two measures are orthogonal. Therefore, in principle, these two measures can provide insights into the effect of these individual differences on speech planning respectively.

Figure 3

Correlation Between WMC Score and Music Perception Score



Individual Difference Results for the 10 Participants Included in the Analysis. Ten participants' data were used in the regression analysis. Table 4 shows the descriptive statistics for the individual difference measures for these 10 participants.

Table 4

Mean and Standard Deviation for the Raw Scores in the Individual Difference Tests for the 10 Participants Included in the Current Analysis

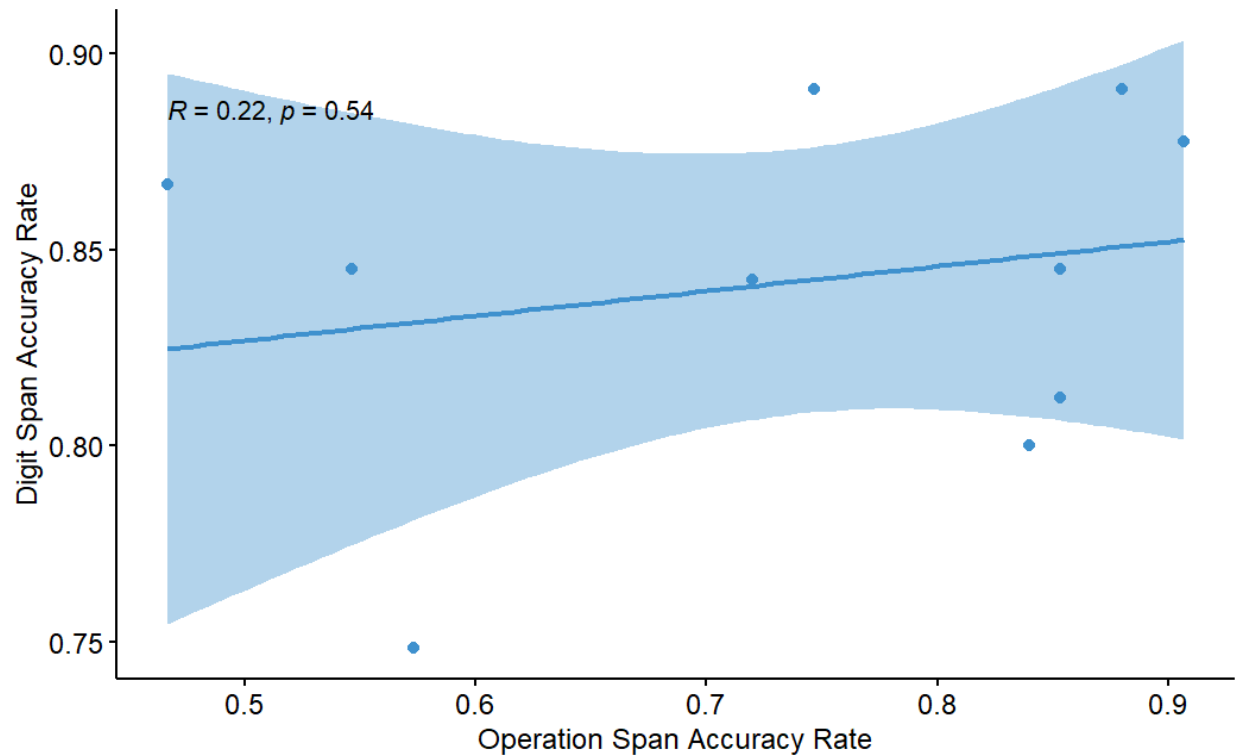
Test	Mean	Standard deviation
Digit Span Test (165)	135.6	9.24
Operation Span Test (75)	55.4	11.82
Rhythmic Contour Test (31)	23.1	4.84
Metric Test (30)	26.7	3.83

Note. The number in parentheses after the test name indicates the highest score a participant might get on this test.

There was a positive but not significant correlation between the 10 participants' performances on the digit span test and the operation span test ($r = 0.22$, $p = 0.54$; Figure 4). Given that the correlation between the accuracy rates for the two WMC tests was significant based on the correlation test carried out with all 58 participants' data, we suspect that the nonsignificant correlation result here is due to the small sample size.

Figure 4

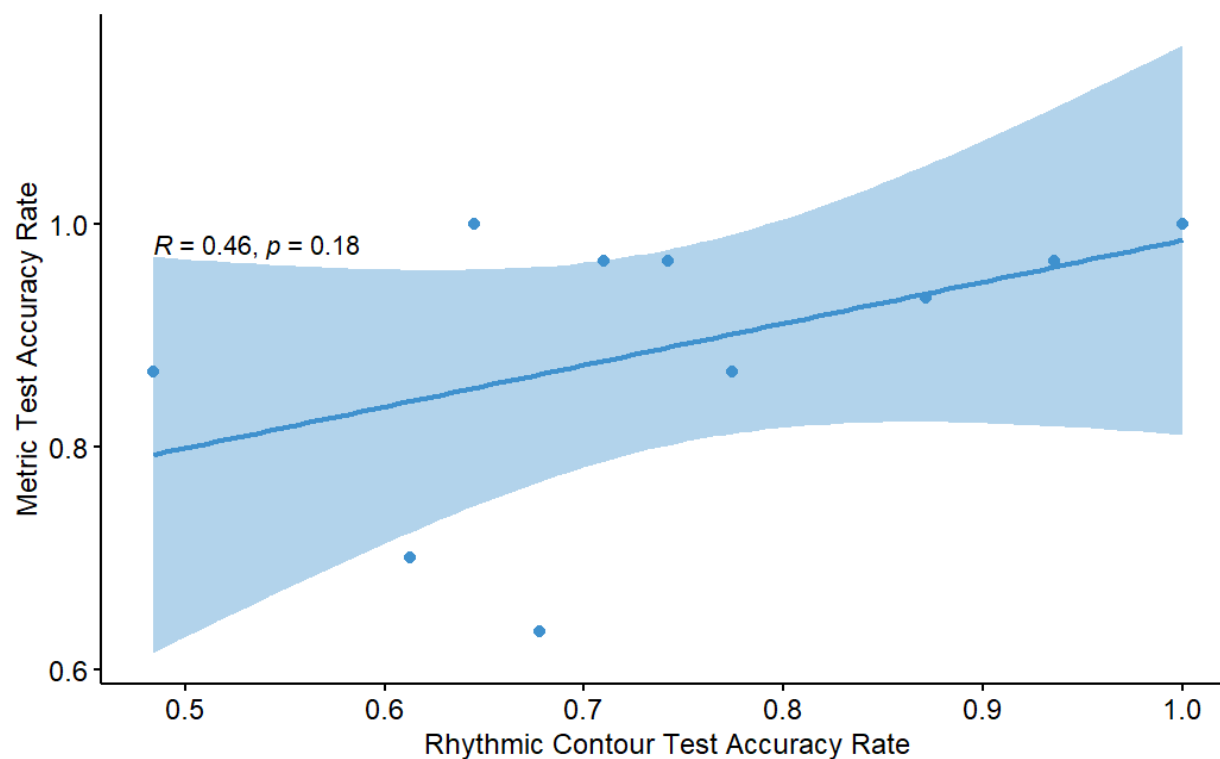
Correlation Between Digit Span Test Accuracy Rate and Operation Span Test Accuracy Rate for the 10 Participants Included in the Current Analysis



There was a positive but not significant correlation between the accuracy rates for the two music perception tests ($r = 0.46, p = 0.18$; Figure 5). Similar to the correlation between the two WMC tests, we suspect that the nonsignificant correlation result is due to the small sample size.

Figure 5

Correlation Between Rhythmic Contour Test Accuracy Rate and Metric Test Accuracy Rate for the 10 Participants Included in the Current Analysis



There was a negative correlation between an individual's composite WMC score and composite music perception score ($r = -0.26, p = 0.48$; Figure 6). This negative correlation was driven by the unbalanced distribution of the WMC and music perception scores within the 10 participants. Table 5 shows the number of participants falling into each category after we median splitted the WMC score and the music perception score. For example, among the 10 participants included in the current analysis, the participants who scored the top five WMC scores were assigned to the high WMC score category. Similarly, among all 58 participants, the participants who scored the top 29 WMC scores were assigned to the high WMC score category.

Figure 6

Correlation Between WMC Score and Music Perception Score for the 10 Participants Included in the Current Analysis

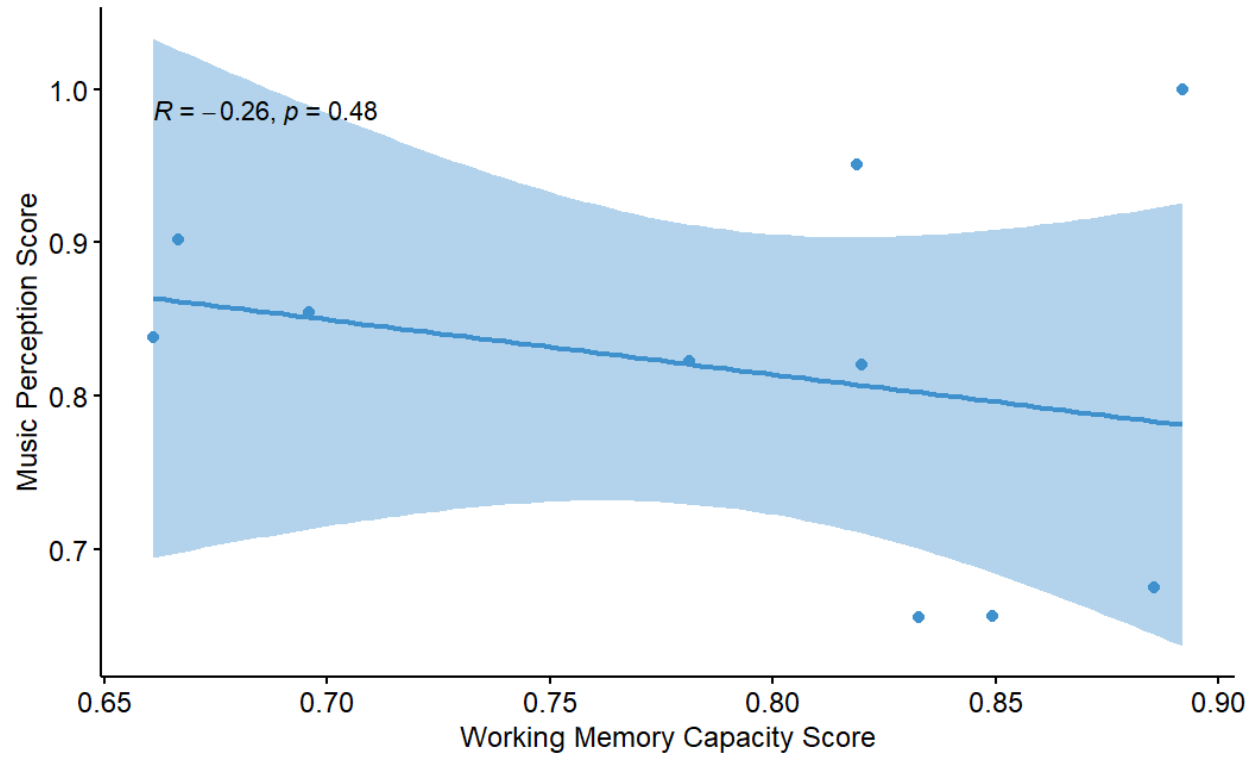


Table 5*Number of Participants with High Versus Low WMC Scores and Music Perception Scores*

		Music Perception Score	
		High	Low
10 participants	WMC Score	High	1
		Low	4
58 participants	WMC Score	High	15
		Low	14

Regression Model Results***Word Pair Repetition Results***

Mean Word Pair Duration: Initial Segmental Overlap Model (Table 6). There was no main effect of metrical overlap. There was a significant main effect of initial stress ($\beta = -16.94, p = 0.006$), such that the mean word pair duration was shorter when the first syllable was stressed than when the first syllable was unstressed (Figure 7 plot A). There was a main effect of initial segmental overlap ($\beta = 32.38, p < 0.001$), such that the mean word pair duration was longer when there was initial segmental overlap than when there was no initial segmental overlap (Figure 7 plot B). There was no interaction effect between metrical overlap and initial stress. There was an interaction effect between metrical overlap and initial segmental overlap ($\beta = 24.78, p = 0.046$), which was mainly driven by the long mean word pair duration when there were both initial segmental and metrical overlaps (Figure 8). There was an interaction effect between initial stress and initial segmental overlap ($\beta = 41.94, p = 0.001$), which was mainly

driven by the short mean word pair duration when there was no initial segmental overlap and the first syllable was stressed (Figure 9). There was a marginally significant three-way interaction between metrical overlap, initial stress, and initial segmental overlap, which was mainly driven by the long duration when there were both metrical and initial segmental overlaps and the first syllable was stressed.

Table 6

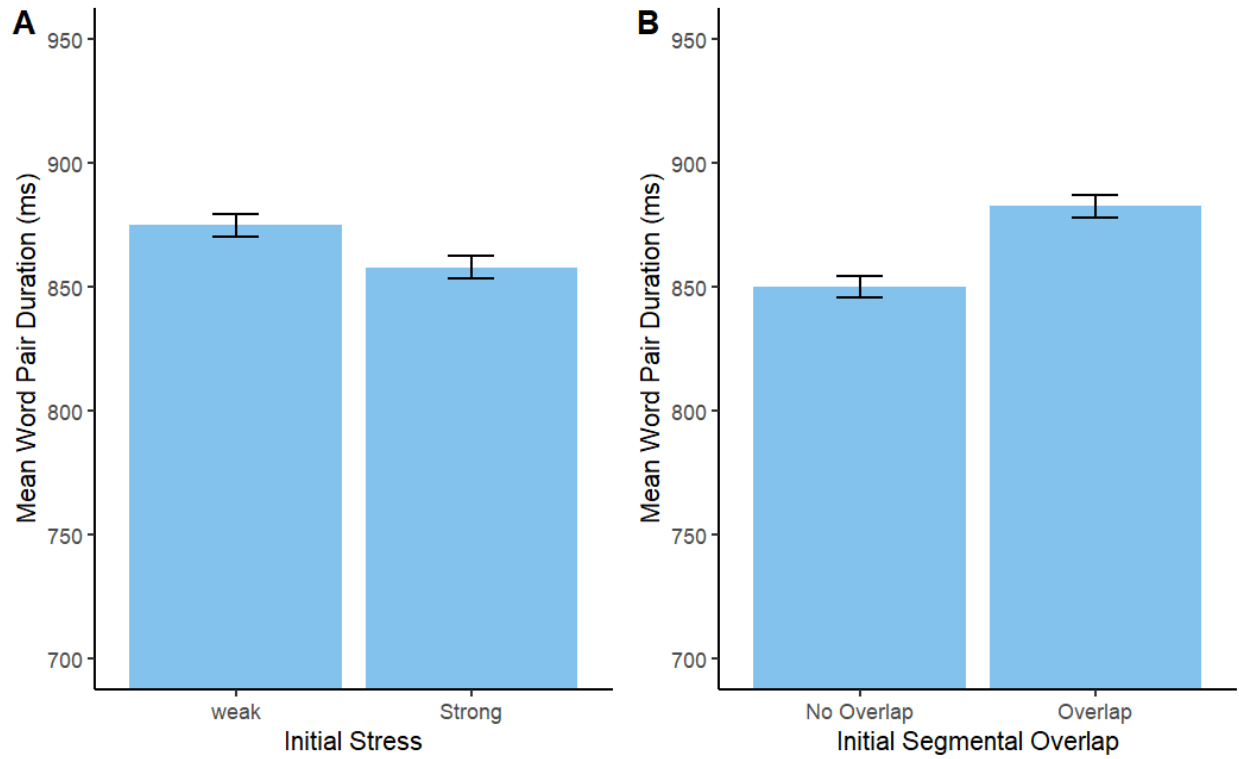
The Initial Segmental Overlap Model for Mean Word Pair Duration

Predictors	Estimates	CI	<i>p</i>
(Intercept)	866.57	[860.47, 872.66]	<0.001
Metrical Overlap	8.77	[-3.41, 20.96]	0.158
Initial Stress	-16.94	[-29.12, -4.75]	0.006
Initial Segmental Overlap	32.38	[20.19, 44.56]	<0.001
Metrical Overlap × Initial Stress	2.16	[-22.21, 26.53]	0.862
Metrical Overlap × Initial Segmental Overlap	24.78	[0.40, 49.15]	0.046
Initial Stress × Initial Segmental Overlap	41.94	[17.57, 66.31]	0.001
Metrical Overlap × Initial Stress × Initial Segmental Overlap	48.04	[-0.71, 96.78]	0.053

Note. $R^2_{\text{adj}} = 0.031$. Significant effects are bolded.

Figure 7

Mean Word Pair Duration by Initial Stress and Initial Segmental Overlap



Note. A: Mean word pair duration by initial stress. B: Mean word pair duration by initial segmental overlap.

Figure 8

Mean Word Pair Duration by Initial Segmental Overlap and Metrical Overlap

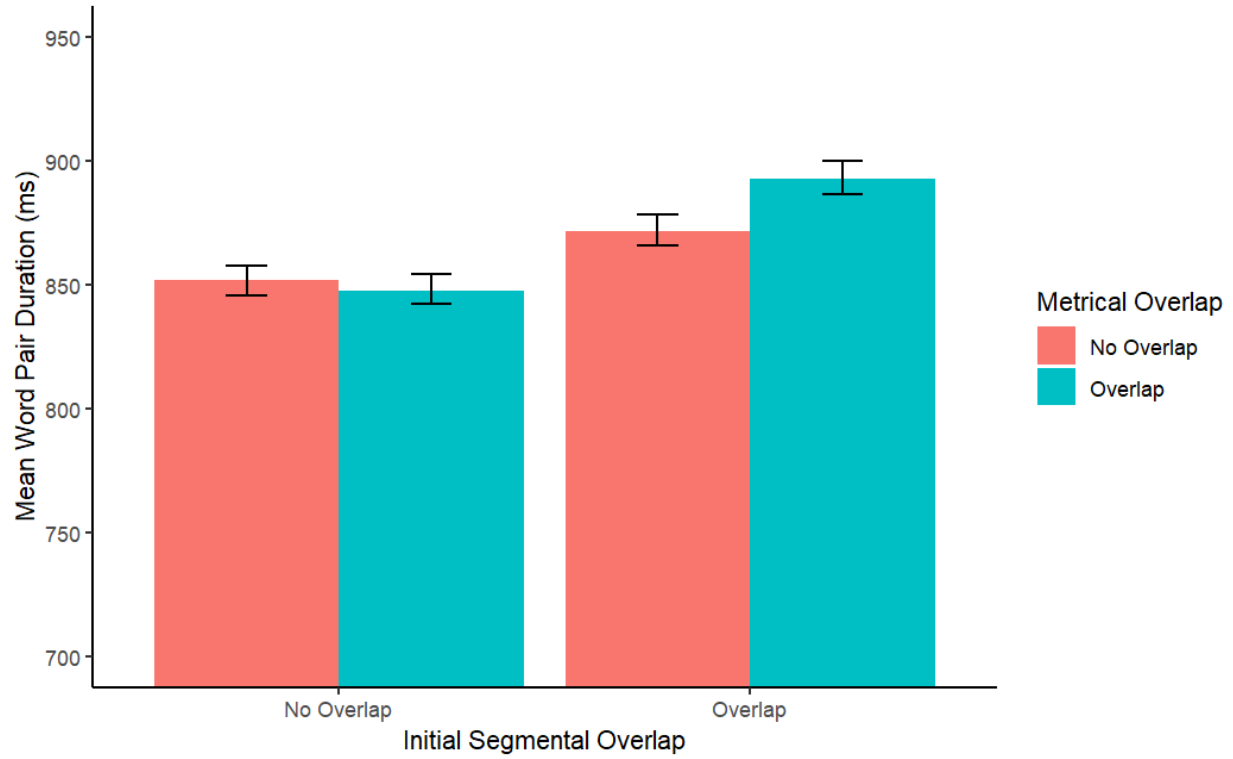
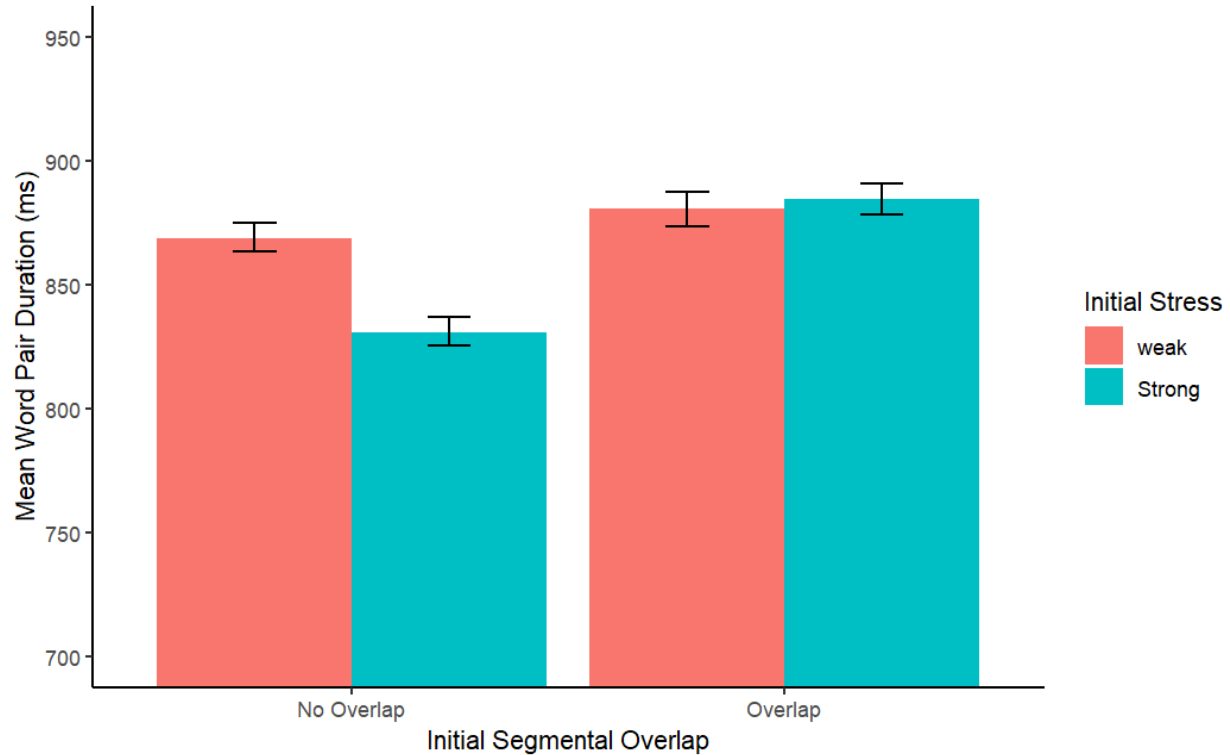


Figure 9

Mean Word Pair Duration by Initial Segmental Overlap and Initial Stress



Mean Word Pair Duration: Phonological Distance Model (Table 7). There was no main effect of metrical overlap. There was a main effect of initial stress ($\beta = -19.32$, $p = 0.003$), such that the mean word pair duration was shorter when the first syllable was stressed than when the first syllable was unstressed. There was a main effect of phonological distance ($\beta = -10.76$, $p < 0.001$; Figure 10), such that the larger the phonological distance was, the shorter the mean word pair duration was. There was no interaction effect between metrical overlap and initial stress. There was no interaction effect between metrical overlap and phonological distance. There was an interaction effect between initial stress and phonological distance ($\beta = -16.04$, $p = 0.001$; Figure 11), such that the larger the phonological distance was, the stronger the effect of

initial stress was. There was no three-way interaction between metrical overlap, initial stress, and phonological distance.

Table 7

The Phonological Distance Model for Mean Word Pair Duration

Predictors	Estimates	CI	<i>p</i>
(Intercept)	865.71	[859.28, 872.14]	<0.001
Metrical Overlap	-0.70	[-13.56, 12.16]	0.915
Initial Stress	-19.32	[-32.18, -6.46]	0.003
Phonological Distance	-10.76	[-15.39, -6.13]	<0.001
Metrical Overlap × Initial Stress	-16.25	[-41.97, 9.47]	0.215
Metrical Overlap × Phonological Distance	-4.63	[-13.88, 4.63]	0.327
Initial Stress × Phonological Distance	-16.04	[-25.30, -6.78]	0.001
Metrical Overlap × Initial Stress × Phonological Distance	-1.81	[-20.32, 16.70]	0.848

Note. $R^2_{\text{adj}} = 0.024$. Significant effects are bolded.

Figure 10

Mean Word Pair Duration by Phonological Distance

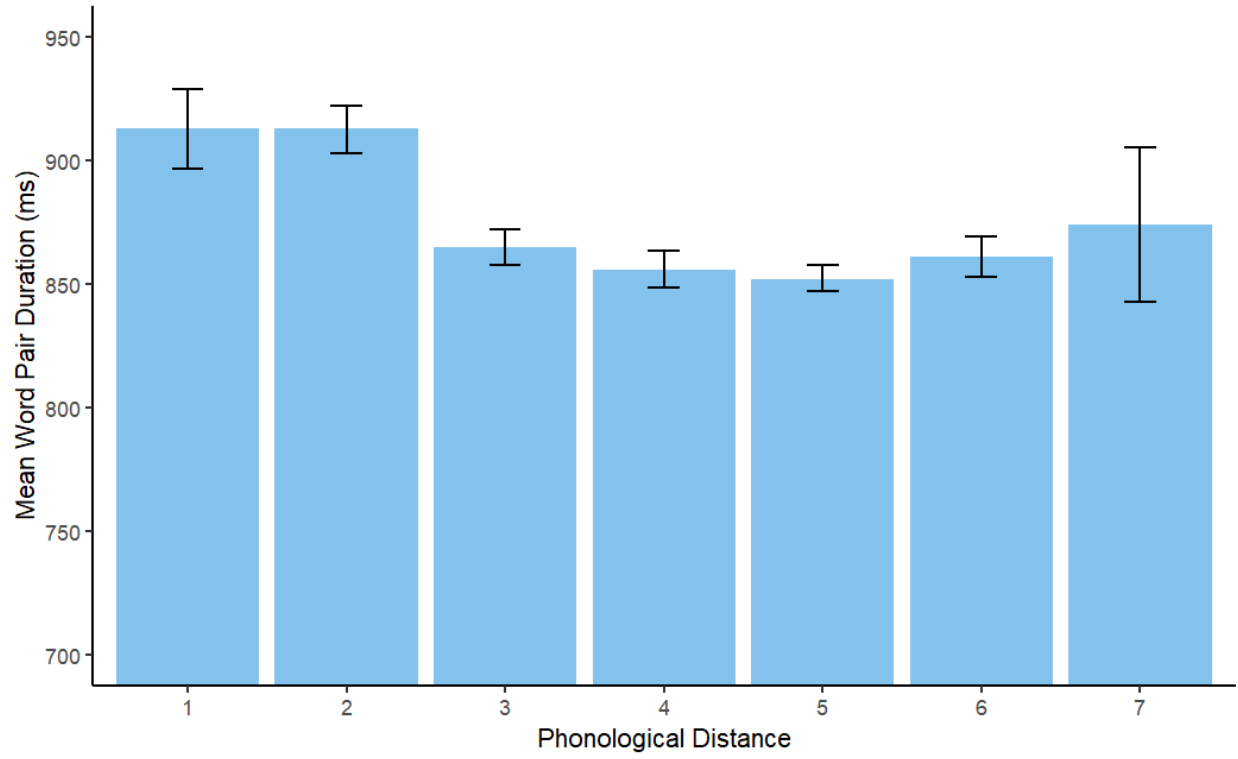
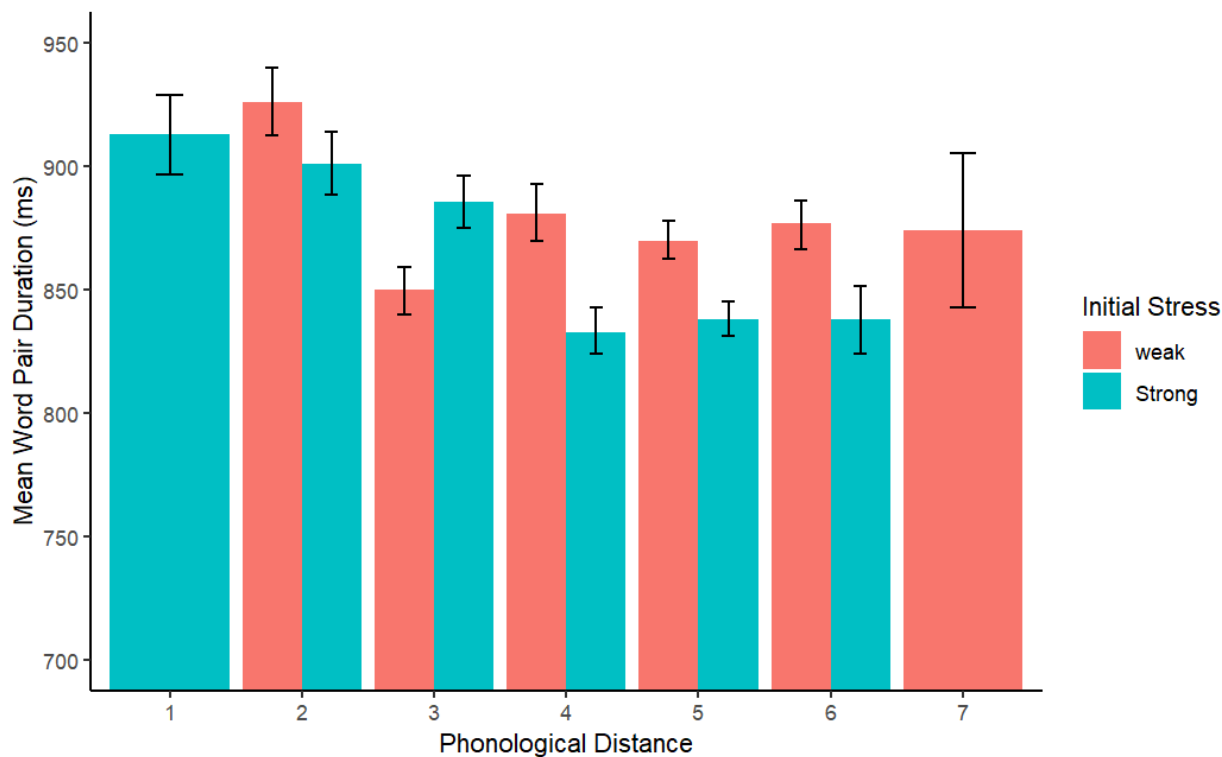


Figure 11

Mean Word Pair Duration by Phonological Distance and Initial Stress



Number of Repetitions: Initial Segmental Overlap Model (Table 8). There was a main effect of metrical overlap ($\beta = -0.25, p = 0.001$), such that there were fewer correct and complete repetitions when there was metrical overlap (Figure 12 plot A). There was a main effect of initial stress ($\beta = 0.20, p = 0.009$), such that there were more repetitions when the first syllable was stressed (Figure 12 plot B). There was a main effect of initial segmental overlap ($\beta = -0.36, p < 0.001$), such that there were fewer repetitions when there was initial segmental overlap (Figure 12 plot C). There was no interaction effect between metrical overlap and initial stress. There was an interaction effect between metrical overlap and initial segmental overlap ($\beta = -0.43, p = 0.006$): There were the fewest repetitions when there were both metrical overlap and initial segmental overlap (Figure 13). There was an interaction effect between initial stress

and initial segmental overlap ($\beta = -0.43, p = 0.005$): There were the most repetitions when there was initial segmental overlap and the first syllable was stressed (Figure 14). There was also a three-way interaction between metrical overlap, initial stress, and initial segmental overlap ($\beta = -0.75, p = 0.015$; Figure 15). T-tests revealed that, first, when there was metrical overlap and the first syllable was stressed, there were more repetitions when there was no initial segmental overlap than when there was initial segmental overlap ($t(383.52) = 6.43, p < 0.001$); second, when there was metrical overlap but not initial segmental overlap, there were more repetitions when the first syllable was stressed than when the first syllable was unstressed ($t(372.11) = -4.01, p < 0.001$). That is, this three-way interaction was mainly driven by the large number of repetitions when there was metrical overlap, no initial segmental overlap, and the first syllable was stressed.

Table 8

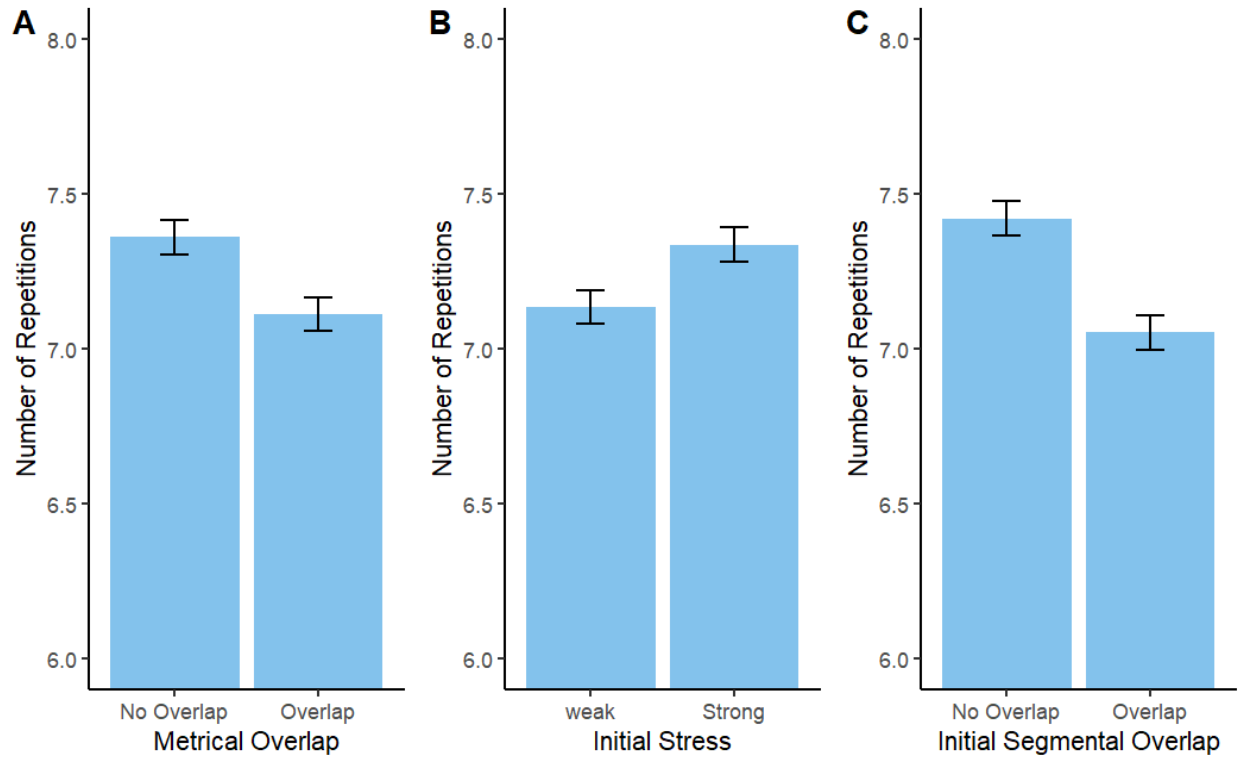
The Initial Segmental Overlap Model for Number of Repetitions

Predictors	Estimates	CI	<i>p</i>
(Intercept)	7.24	[7.16, 7.31]	<0.001
Metrical Overlap	-0.25	[-0.40, -0.10]	0.001
Initial Stress	0.20	[0.05, 0.35]	0.009
Initial Segmental Overlap	-0.36	[-0.52, -0.21]	<0.001
Metrical Overlap × Initial Stress	-0.03	[-0.33, 0.27]	0.850
Metrical Overlap × Initial Segmental Overlap	-0.43	[-0.73, -0.13]	0.006
Initial Stress × Initial Segmental Overlap	-0.43	[-0.73, -0.13]	0.005
Metrical Overlap × Initial Stress × Initial Segmental Overlap	-0.75	[-1.35, -0.15]	0.015

Note. $R^2_{\text{adj}} = 0.034$. Significant effects are bolded.

Figure 12

Number of Repetitions by Metrical Overlap, by Initial Stress, and by Initial Segmental Overlap



Note. A: Number of repetitions by metrical overlap. B: Number of repetitions by initial stress. C: Number of repetitions by initial segmental overlap.

Figure 13

Number of Repetitions by Initial Segmental Overlap and Metrical Overlap

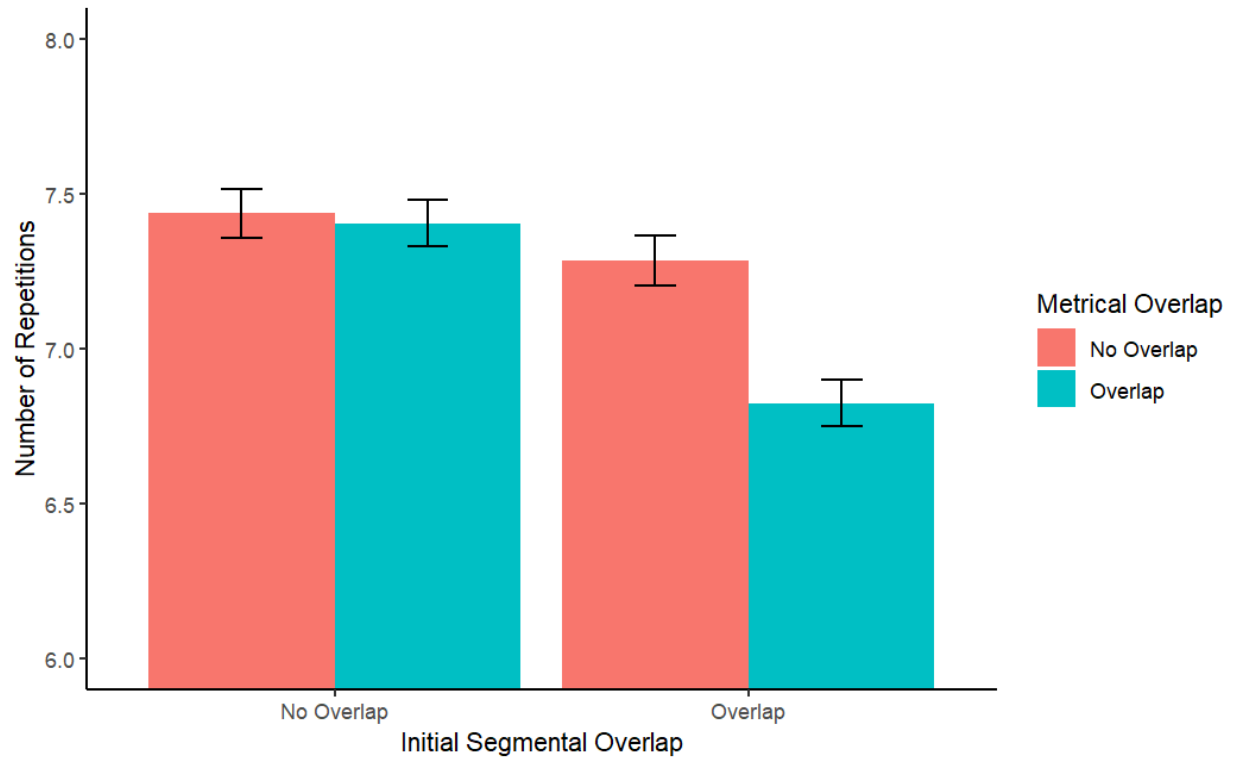


Figure 14

Number of Repetitions by Initial Segmental Overlap and Initial Stress

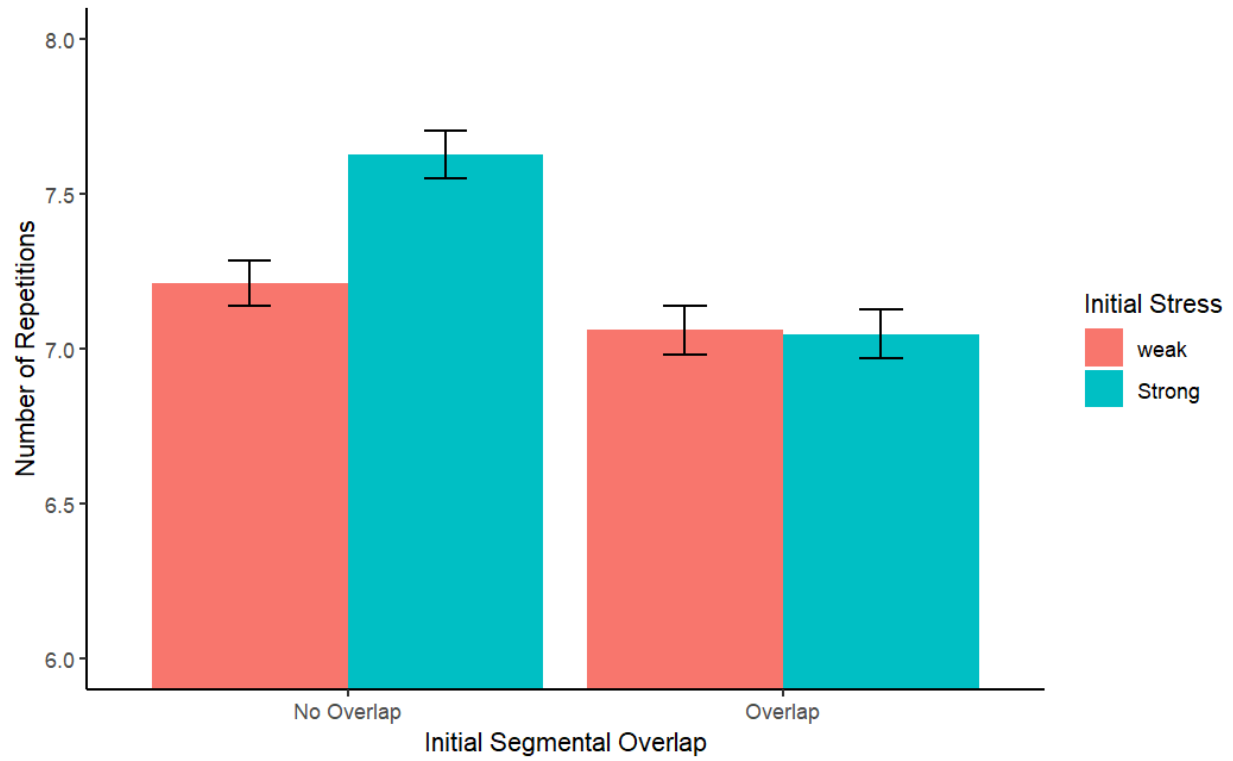
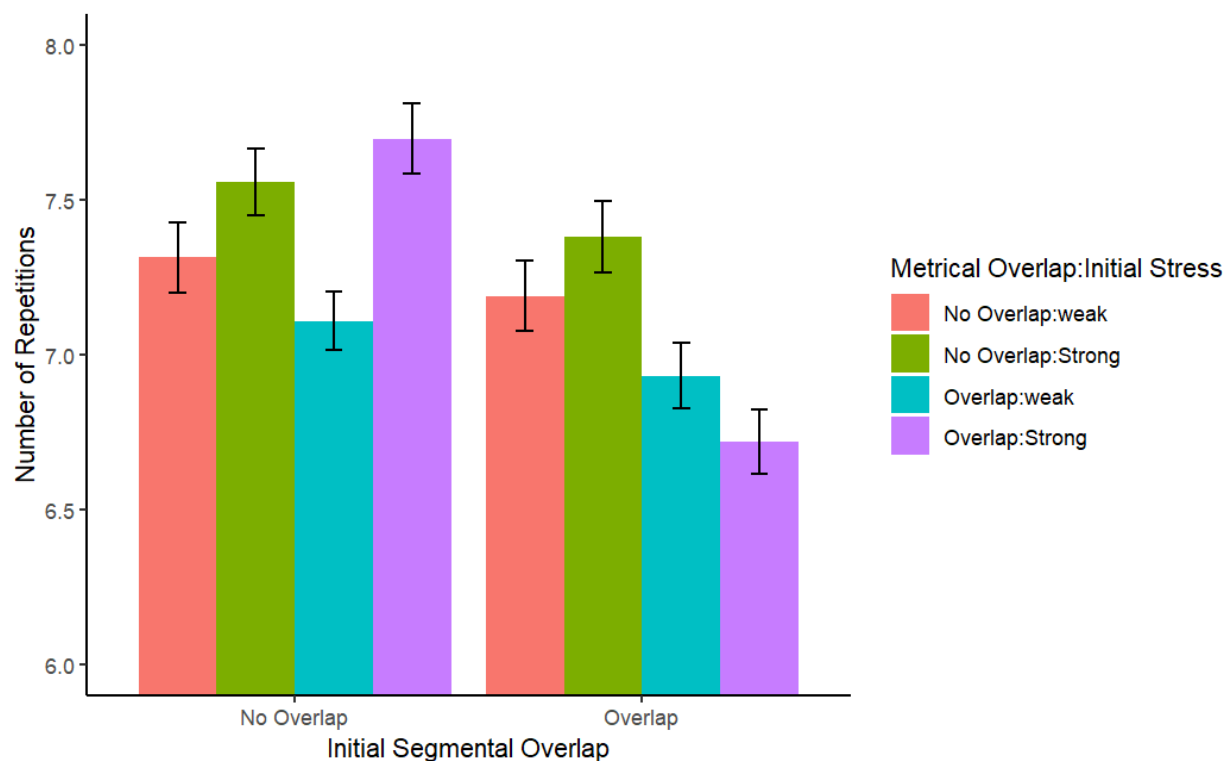


Figure 15

Number of Repetitions by Initial Segmental Overlap, Metrical Overlap, and Initial stress



Number of Repetitions: Phonological Distance Model (Table 9). There was no main effect of metrical overlap. There was a main effect of initial stress ($\beta = 0.26, p = 0.001$), such that there were more repetitions when the first syllable was stressed. There was a main effect of phonological distance ($\beta = 0.16, p < 0.001$), such that the larger the phonological distance was, the more repetitions there were (Figure 16). There was no interaction effect between metrical overlap and initial stress, nor interaction effect between metrical overlap and phonological distance. There was a significant interaction effect between initial stress and phonological distance ($\beta = 0.14, p = 0.017$), such that the larger the phonological distance was, the stronger the effect of initial stress was (Figure 17). There was no three-way interaction between metrical overlap, initial stress, and phonological distance.

Table 9*The Phonological Distance Model for Number of Repetitions*

Predictors	Estimates	CI	<i>p</i>
(Intercept)	7.25	[7.17, 7.32]	<0.001
Metrical Overlap	-0.12	[-0.27, 0.04]	0.154
Initial Stress	0.26	[0.10, 0.42]	0.001
Phonological Distance	0.16	[0.10, 0.21]	<0.001
Metrical Overlap × Initial Stress	0.16	[-0.15, 0.48]	0.309
Metrical Overlap × Phonological Distance	0.06	[-0.05, 0.18]	0.287
Initial Stress × Phonological Distance	0.14	[0.02, 0.25]	0.017
Metrical Overlap × Initial Stress × Phonological Distance	0.09	[-0.14, 0.32]	0.442

Note. $R^2_{\text{adj}} = 0.032$. Significant effects are bolded.

Figure 16

Number of Repetitions by Phonological Distance

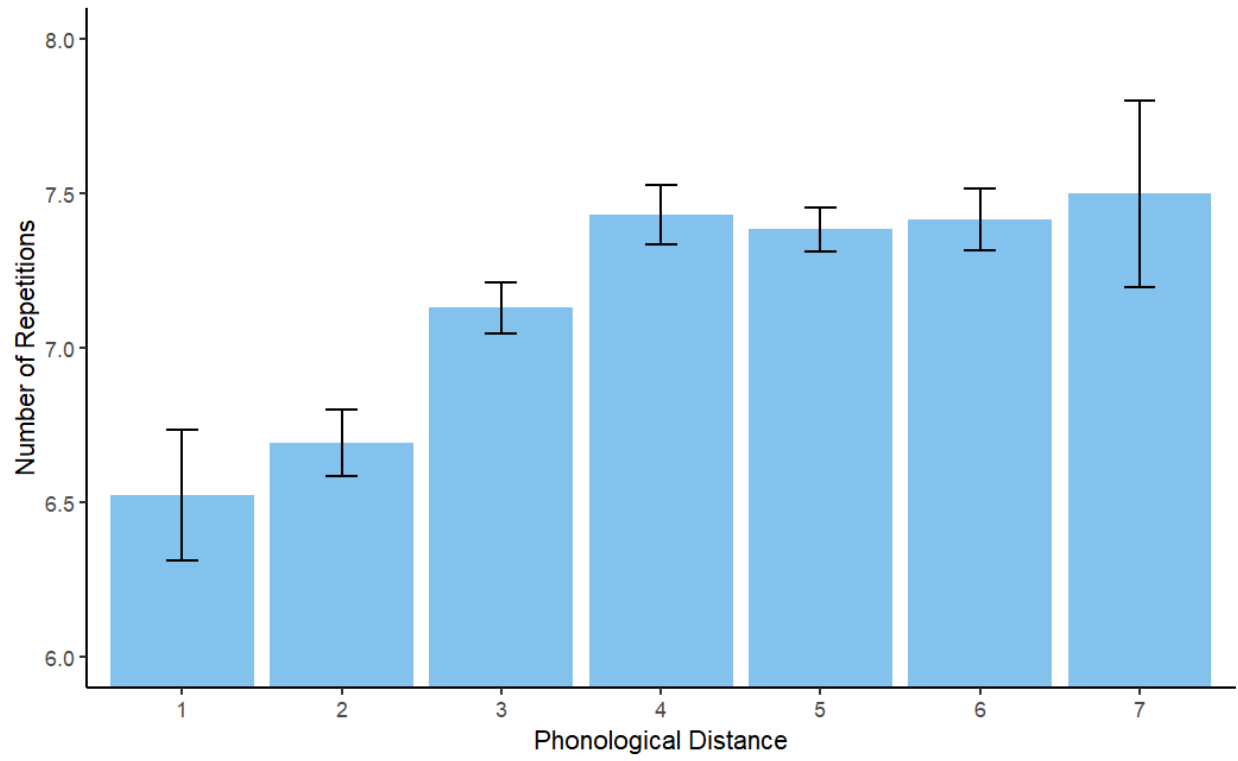
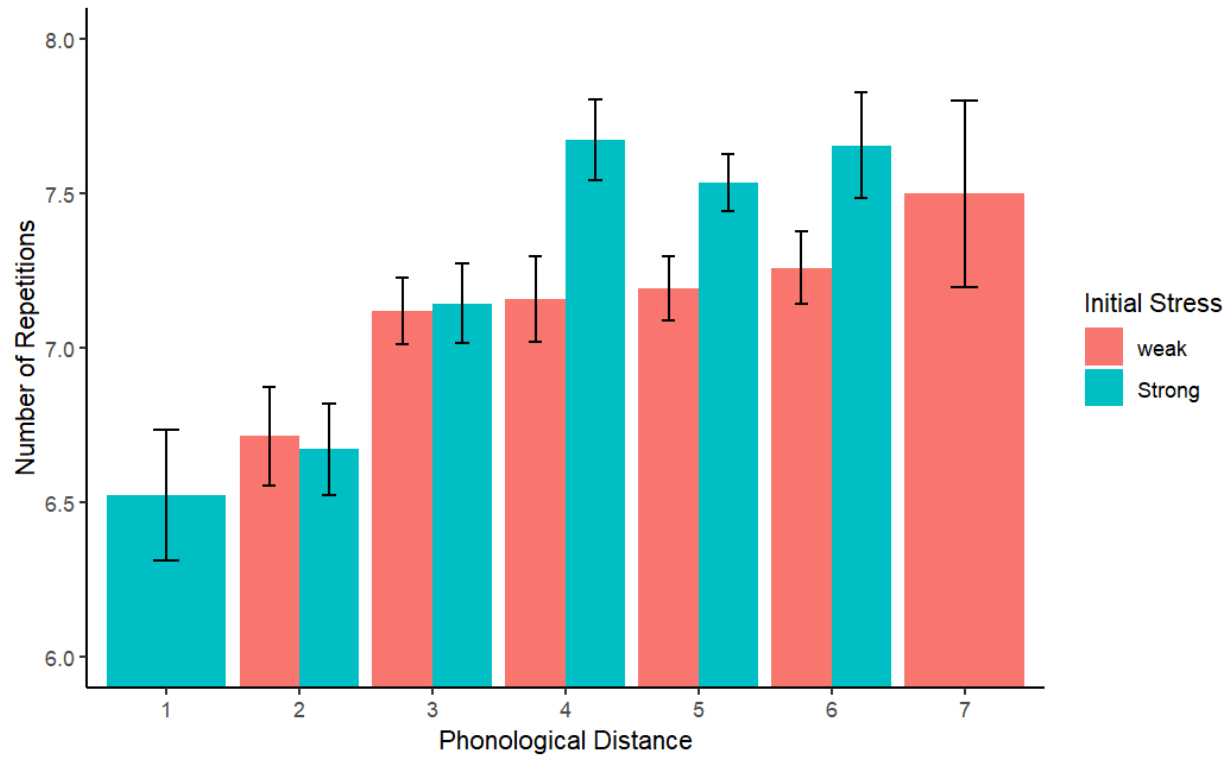


Figure 17

Number of Repetitions by Phonological Distance and Initial Stress



Number of Errors: Initial Segmental Overlap Model (Table 10). There was no significant predictor for the number of errors in the initial segmental overlap model.

Table 10*The Initial Segmental Overlap Model for Number of Errors*

Predictors	Estimates	CI	<i>p</i>
(Intercept)	0.68	[0.63, 0.74]	<0.001
Metrical Overlap	0.09	[-0.02, 0.21]	0.104
Initial Stress	-0.01	[-0.12, 0.11]	0.900
Initial Segmental Overlap	0.03	[-0.08, 0.14]	0.578
Metrical Overlap × Initial Stress	0.11	[-0.12, 0.33]	0.358
Metrical Overlap × Initial Segmental Overlap	0.03	[-0.19, 0.26]	0.771
Initial Stress × Initial Segmental Overlap	0.07	[-0.16, 0.29]	0.563
Metrical Overlap × Initial Stress × Initial Segmental Overlap	0.06	[-0.39, 0.51]	0.800

Note. $R^2_{\text{adj}} = -0.002$. Significant effects are bolded.

Number of Errors: Phonological Distance Model (Table 11). There was no significant predictor for the number of errors in the phonological distance model.

Table 11*The Phonological Distance Model for Number of Errors*

Predictors	Estimates	CI	<i>p</i>
(Intercept)	0.69	[0.63, 0.75]	<0.001
Metrical Overlap	0.07	[-0.05, 0.18]	0.280
Initial Stress	-0.02	[-0.13, 0.10]	0.794
Phonological Distance	-0.03	[-0.08, 0.01]	0.110
Metrical Overlap × Initial Stress	0.10	[-0.14, 0.33]	0.427
Metrical Overlap × Phonological Distance	0.03	[-0.06, 0.11]	0.549
Initial Stress × Phonological Distance	0.00	[-0.08, 0.09]	0.984
Metrical Overlap × Initial Stress × Phonological Distance	-0.01	[-0.18, 0.16]	0.888

Note. $R^2_{\text{adj}} = -0.000$. Significant effects are bolded.

WMC Results

Mean Word Pair Duration: WMC Model (Table 12). All significant predictors in the initial segmental overlap model for mean word pair duration remained significant. In addition, there was a main effect of WMC score ($\beta = 62.06$, $p < 0.001$), such that the mean word pair duration was longer for people with better WMC scores (Figure 18). While the three-way interaction effect between metrical overlap, initial stress, and initial segmental overlap was only marginally significant in the initial segmental overlap model for mean word pair duration, it became significant in this model ($\beta = 48.36$, $p = 0.045$). T-test revealed that when there was initial segmental overlap and the first syllable was stressed, the mean word pair duration was longer when there was metrical overlap than when there was no metrical overlap ($t(387.67) = -2.78$, $p = 0.006$; Figure 19).

Table 12*The WMC Score Model for Mean Word Pair Duration*

Predictor	Estimates	CI	<i>p</i>
(Intercept)	866.52	[860.62, 872.42]	<0.001
Metrical Overlap	9.26	[-2.54, 21.06]	0.124
Initial Stress	-17.49	[-29.29, -5.70]	0.004
Initial Segmental Overlap	32.38	[20.58, 44.17]	<0.001
WMC Score	62.06	[50.26, 73.86]	<0.001
Metrical Overlap × Initial Stress	1.56	[-22.04, 25.16]	0.897
Metrical Overlap × Initial Segmental Overlap	24.97	[1.37, 48.56]	0.038
Initial Stress × Initial Segmental Overlap	41.25	[17.65, 64.85]	0.001
Metrical Overlap × WMC Score	-4.42	[-28.02, 19.18]	0.713
Initial Segmental Overlap × WMC Score	4.97	[-18.63, 28.56]	0.680
Metrical Overlap × Initial Stress × Initial Segmental Overlap	48.36	[1.17, 95.56]	0.045
Metrical Overlap × Initial Segmental Overlap × WMC Score	1.40	[-45.80, 48.60]	0.954

Note. $R^2_{\text{adj}} = 0.091$. Significant effects are bolded.

Figure 18

Mean Word Pair Duration by WMC Score

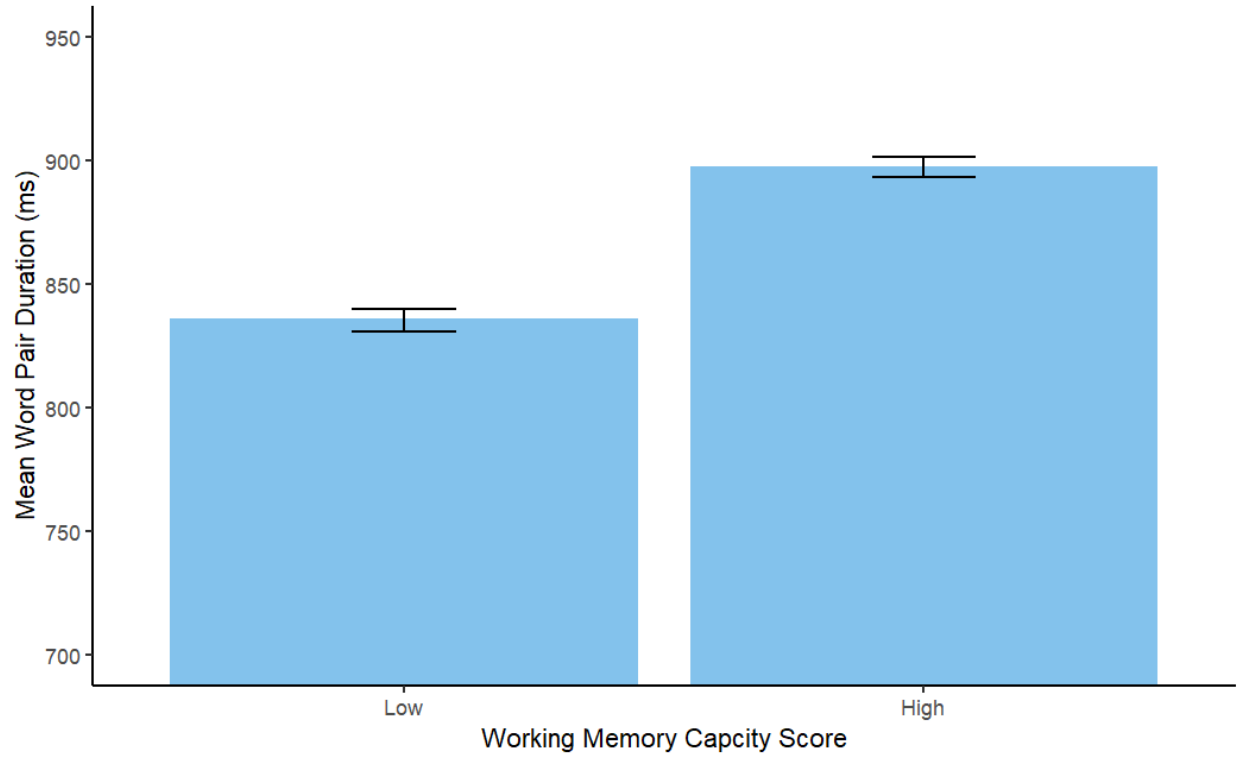
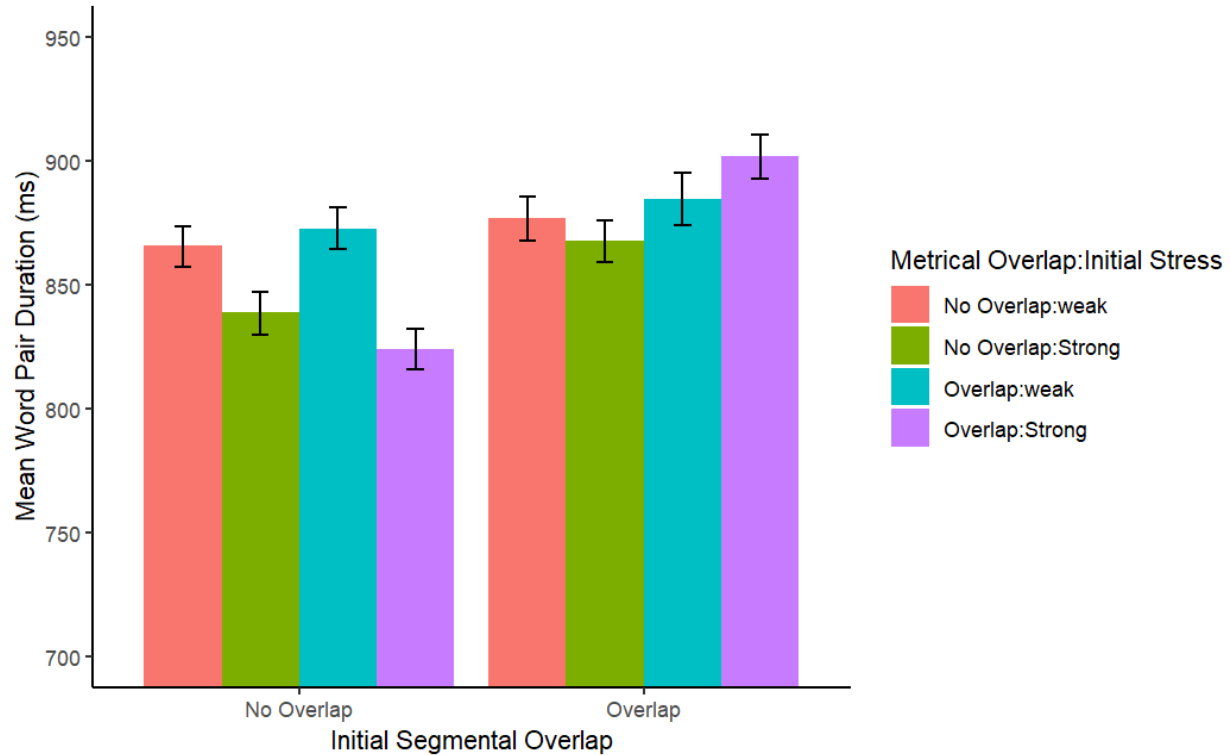


Figure 19

Mean Word Pair Duration by Initial Segmental Overlap, Metrical Overlap, and Initial Stress



Number of Repetitions: WMC Model (Table 13). All significant predictors in the initial segmental overlap model for the number of repetitions remained significant. In addition, there was a main effect of WMC score, such that people with higher WMC score had fewer repetitions (Figure 20). There was no interaction effect between WMC score and other variables.

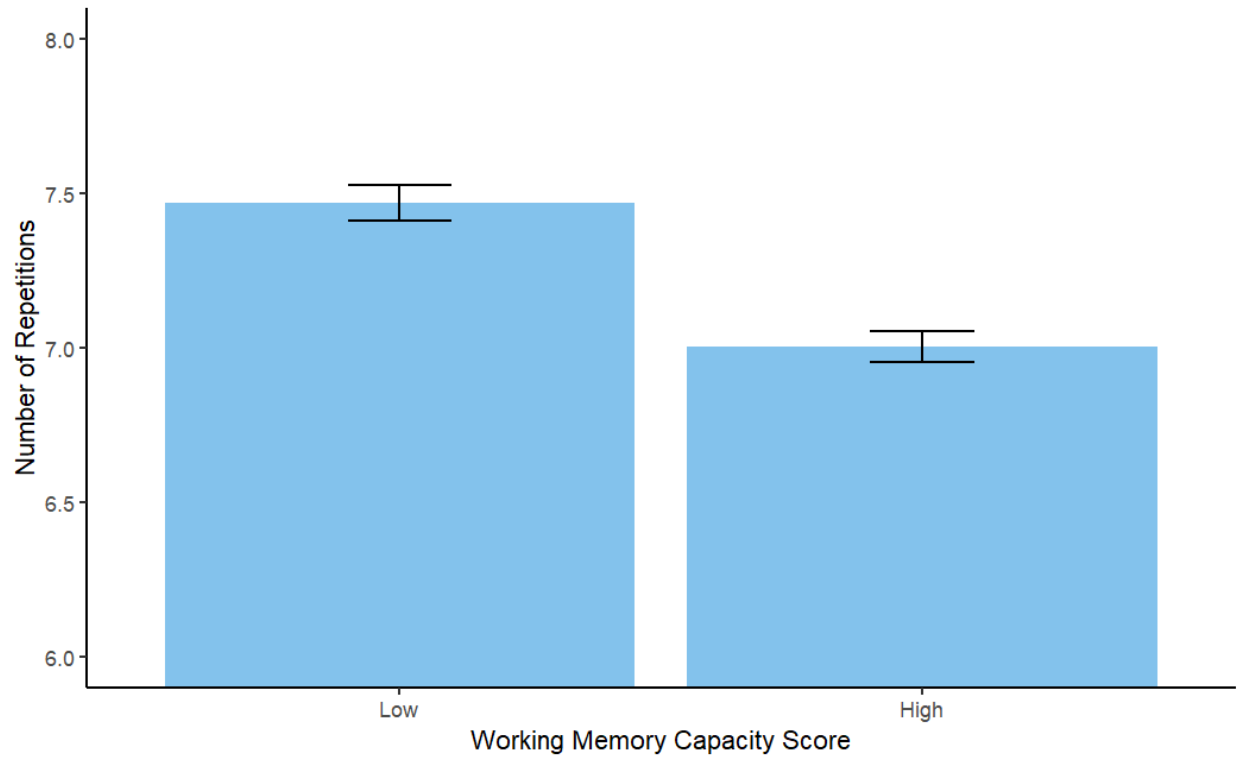
Table 13*The WMC Score Model for Number of Repetitions*

Predictor	Estimates	CI	<i>p</i>
(Intercept)	7.24	[7.16, 7.31]	<0.001
Metrical Overlap	-0.25	[-0.40, -0.10]	0.001
Initial Stress	0.21	[0.06, 0.35]	0.007
Initial Segmental Overlap	-0.37	[-0.51, -0.22]	<0.001
WMC Score	-0.47	[-0.62, -0.32]	<0.001
Metrical Overlap × Initial Stress	-0.03	[-0.32, 0.27]	0.865
Metrical Overlap × Initial Segmental Overlap	-0.43	[-0.73, -0.13]	0.005
Initial Stress × Initial Segmental Overlap	-0.42	[-0.72, -0.12]	0.005
Metrical Overlap × WMC Score	0.24	[-0.06, 0.53]	0.121
Initial Segmental Overlap × WMC Score	0.03	[-0.27, 0.33]	0.833
Metrical Overlap × Initial Stress × Initial Segmental Overlap	-0.75	[-1.35, -0.16]	0.013
Metrical Overlap × Initial Segmental Overlap × WMC Score	-0.21	[-0.81, 0.39]	0.488

Note. $R^2_{\text{adj}} = 0.057$. Significant effects are bolded.

Figure 20

Number of Repetitions by WMC Score

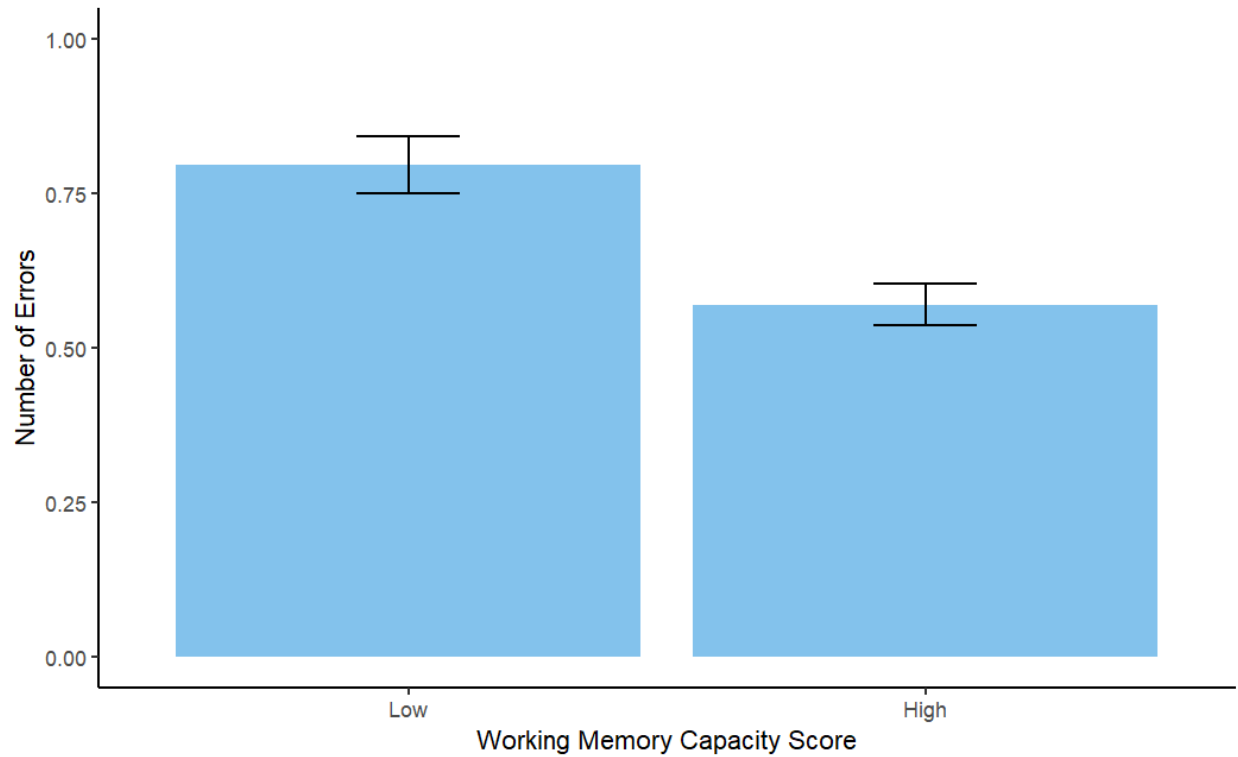


Number of Errors: WMC Model (Table 14). There was a main effect of WMC score, such that people with higher WMC scores made fewer errors than people with lower WMC scores (Figure 21). There was no other significant predictor in the model.

Table 14*The WMC Score Model for Number of Errors*

Predictor	Estimates	CI	<i>p</i>
(Intercept)	0.68	[0.63, 0.74]	<0.001
Metrical Overlap	0.09	[-0.02, 0.20]	0.109
Initial Stress	-0.00	[-0.12, 0.11]	0.932
Initial Segmental Overlap	0.03	[-0.08, 0.14]	0.572
WMC Score	-0.23	[-0.34, -0.11]	<0.001
Metrical Overlap × Initial Stress	0.11	[-0.12, 0.33]	0.348
Metrical Overlap × Initial Segmental Overlap	0.03	[-0.19, 0.26]	0.779
Initial Stress × Initial Segmental Overlap	0.07	[-0.15, 0.29]	0.545
Metrical Overlap × WMC Score	0.02	[-0.21, 0.24]	0.894
Initial Segmental Overlap × WMC Score	-0.09	[-0.31, 0.13]	0.429
Metrical Overlap × Initial Stress × Initial Segmental Overlap	0.06	[-0.39, 0.50]	0.805
Metrical Overlap × Initial Segmental Overlap × WMC Score	0.13	[-0.32, 0.58]	0.564

Note. $R^2_{\text{adj}} = 0.006$. Significant effects are bolded.

Figure 21*Number of Errors by WMC Score****Music Perception Skill Results***

Mean Word Pair Duration: Music Perception Skill Model (Table 15). All significant predictors in the initial segmental overlap model for mean word pair duration remained significant. In addition, there was a main effect of music perception score: People with higher music perception scores had shorter mean word pair durations than people with lower music perception scores (Figure 22). There was no interaction effect between music perception score and metrical overlap.

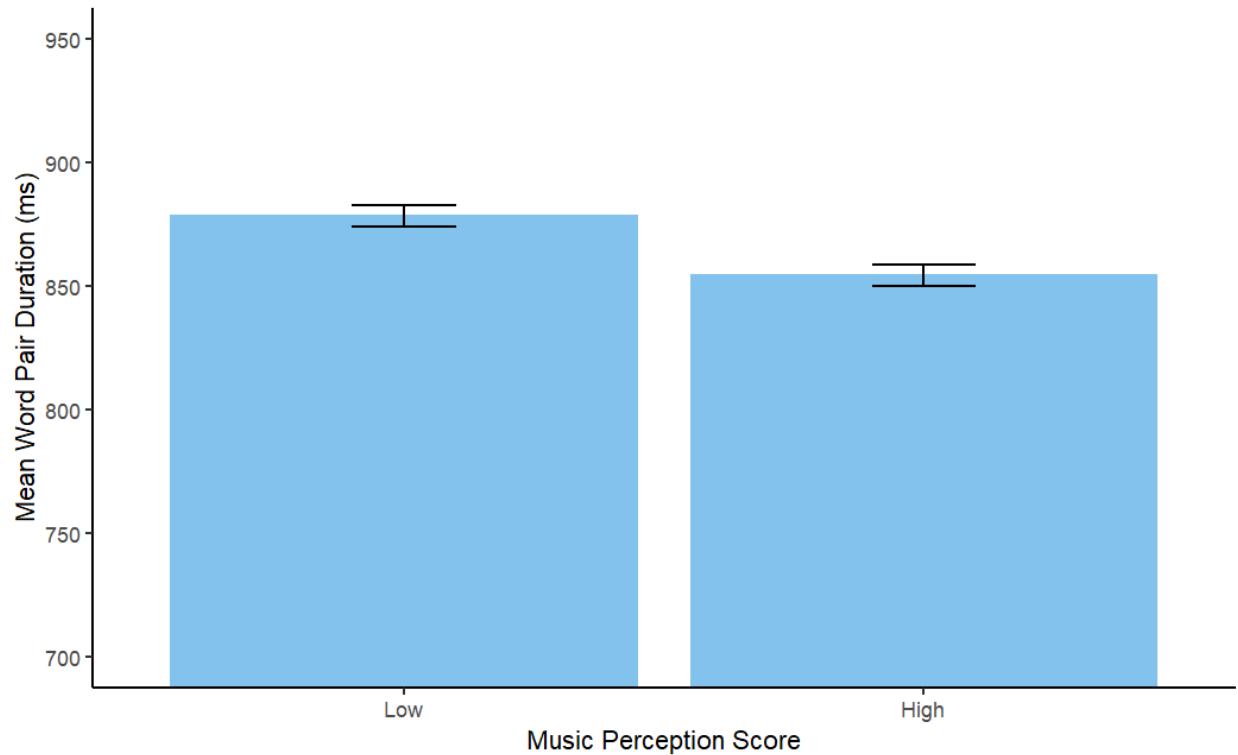
Table 15*The Music Perception Skill Model for Mean Word Pair Duration*

Predictor	Estimates	CI	<i>p</i>
(Intercept)	866.55	[860.48, 872.62]	<0.001
Metrical Overlap	8.92	[-3.21, 21.06]	0.149
Initial Stress	-17.06	[-29.20, -4.93]	0.006
Initial Segmental Overlap	32.49	[20.35, 44.62]	<0.001
Music Perception Score	-24.06	[-36.20, -11.93]	<0.001
Metrical Overlap × Initial Stress	2.08	[-22.19, 26.34]	0.867
Metrical Overlap × Initial Segmental Overlap	25.04	[0.77, 49.31]	0.043
Initial Stress × Initial Segmental Overlap	41.63	[17.36, 65.89]	0.001
Metrical Overlap × Music Perception Score	-5.02	[-29.29, 19.24]	0.685
Metrical Overlap × Initial Stress × Initial Segmental Overlap	47.98	[-0.56, 96.51]	0.053

Note. $R^2_{\text{adj}} = 0.039$. Significant effects are bolded.

Figure 22

Mean Word Pair Duration by Music Perception Score



Number of Repetitions: Music Perception Skill Model. All significant predictors in the initial segmental overlap model for the number of repetitions remained significant. However, there was no main effect of music perception score nor interaction effect between music perception score and metrical overlap.

Number of Errors: Music Perception Skill Model (Table 16). There was a significant main effect of music perception score: People with higher music perception scores produced more speech errors than people with lower music perception scores (Figure 23). There was no interaction effect between music perception score and metrical overlap.

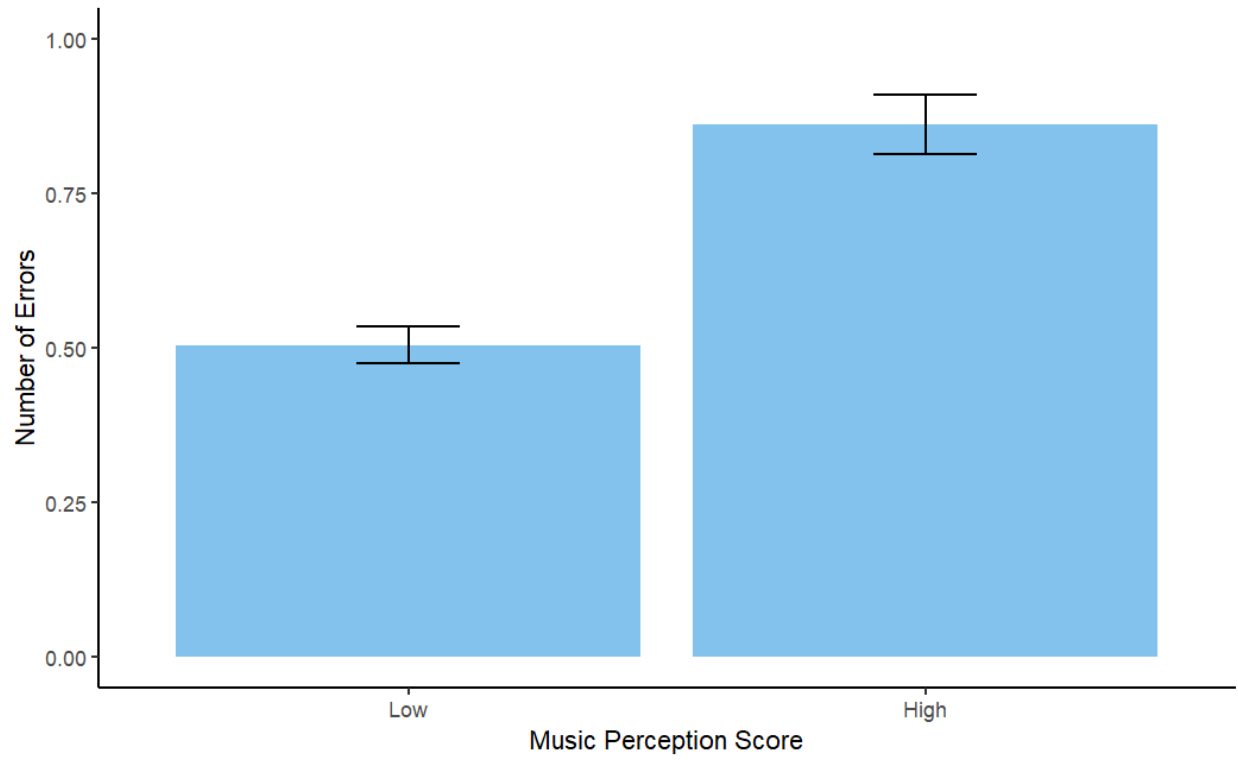
Table 16*The Music Perception Skill Model for Number of Errors*

Predictor	Estimates	CI	<i>p</i>
(Intercept)	0.68	[0.63, 0.74]	<0.001
Metrical Overlap	0.09	[-0.02, 0.20]	0.109
Initial Stress	-0.01	[-0.12, 0.11]	0.924
Initial Segmental Overlap	0.03	[-0.08, 0.14]	0.589
Music Perception Score	0.36	[0.25, 0.47]	<0.001
Metrical Overlap × Initial Stress	0.11	[-0.12, 0.33]	0.349
Metrical Overlap × Initial Segmental Overlap	0.03	[-0.19, 0.25]	0.792
Initial Stress × Initial Segmental Overlap	0.07	[-0.15, 0.29]	0.531
Metrical Overlap × Music Perception Score	-0.03	[-0.26, 0.19]	0.769
Metrical Overlap × Initial Stress × Initial Segmental Overlap	0.06	[-0.39, 0.50]	0.800

Note. $R^2_{\text{adj}} = 0.022$. Significant effects are bolded.

Figure 23

Number of Errors by Music Perception Score



DISCUSSION

The current study aimed to explore whether metrical structure and segmental structure are independently or jointly planned in speech production, and whether individual differences in WMC and music perception skills moderate the relationship between word pair characteristics and speech production difficulty. Specifically, we hypothesized that (1) if metrical structure and segmental structure are independently planned, then speech production should be easier when there is metrical overlap or phonological overlap, which would be shown in shorter mean word pair duration, more correct and complete repetitions, and fewer speech errors; if metrical structure and segmental structure are jointly planned, then speech production should be more challenging when there is metrical overlap and phonological overlap, which would be shown in longer mean word pair duration, fewer correct and complete repetitions, and more speech errors; (2) metrical overlap and phonological overlap would have a smaller effect on speech planning for people with better WMC, as shown by an interaction effect between WMC and metrical overlap, and an interaction effect between WMC and initial segmental overlap; (3) Metrical overlap would have a stronger effect on speech planning for people with better music perception skill, as shown by an interaction effect between music perception skill and metrical overlap. Our results suggest that (1) metrical structure and segmental structure are jointly planned, (2) people with better WMC are worse at speech planning but this underperformance does not depend on the phonological similarity between the two words, and (3) people with better music perception

skills are better at speech planning but this outperformance does not depend on whether the two words in a word pair have metrical overlap or not.

Lexical Representation

Initial Segmental Overlap Models

We found evidence for a joint planning of metrical structure and segmental structure, replicating and extending the findings of Myers and Watson (2021). First, the mean word pair duration was longer when there was initial segmental overlap and when there were both metrical and initial segmental overlaps. There were also fewer repetitions when there was only metrical overlap or only initial segmental overlap, and when there were both metrical and initial segmental overlaps. Moreover, there were the most repetitions when there was metrical overlap, no segmental overlap, and the first syllable was stressed, further suggesting the intertwined relationship between metrical structure and segmental structure. In other words, not only does segmental or metrical overlap matter, but the stress status of the first syllable also plays a role.

However, the mean word pair duration was not lengthened when there was only metrical overlap. This pattern is consistent with the findings in Myers and Watson (2019). Specifically, in an event describing experiment, they found that the word duration of the target word was lengthened only when the prime word and the target word shared the initial segments or shared both metrical structure and initial segments, but not when they only shared metrical structure. Nevertheless, it is also possible that the lack of main effect of metrical overlap on mean word pair duration was due to a weak statistical power: Only 10 participants' data were included in the current analysis. Therefore, we would need to include more data to determine if this is a potentially meaningful null result or merely a lack of statistical power.

In addition, the mean word pair duration was longer and there were fewer repetitions when the first syllable was unstressed, which is in line with Aichert et al. (2016). This effect was expected based on the literature showing that there are more English words following a Strong-weak stress pattern than English words following a weak-Strong stress pattern (Cutler & Carter, 1987). There was also an interaction effect between initial segmental overlap and initial stress, such that the facilitatory effect of having a stressed initial syllable was stronger when there was initial segmental overlap than when there was no initial segmental overlap.

There was no significant effect of any word pair characteristics on the number of errors. One possible reason was the weak statistical power. Another possible reason was that there were a few errors in general. Only 603 out of 1548 trials included in the analysis had at least one error. For the 603 trials with errors, the mean number of errors was only 1.75.

While we found a significant effect of initial segmental overlap on speech planning difficulty, it has to be noted that we did not control how many initial segments are shared between the two words in a word pair. Among the 80 word pairs that have initial segmental overlap, 18 of them share the first phoneme (e.g., “relish relax”). The other 62 word pairs share at least the first two phonemes (e.g., “arrow error”) and at most the first four phonemes (e.g., “ballot ballad”). Therefore, there can be some between-item variance in the effect of initial segmental overlap that is not accounted for in the current analysis. However, because the current study aimed to replicate and extend the results from Myers and Watson (2021), this design limitation was necessary to ensure that our experiment is as similar to the original experiment as possible. Therefore, we decided not to make any changes to the word pairs. Future analysis using a larger sample size should consider including random effects over items to account for these between-item variances.

Phonological Distance Models

The results from the phonological distance models are also consistent with the account that metrical structure is jointly planned with segmental structure. When the phonological distance increased, the mean word pair duration decreased and the number of repetitions increased. This is consistent with the main effect of initial segmental overlap: The speech planning is easier when the two words are less phonologically similar. As in the initial segmental overlap models, when the first syllable was unstressed, the mean word pair duration was longer and there were fewer repetitions. The effect of having a strong initial syllable became stronger as the phonological distance increased, which is similar to the interaction effect between initial segmental overlap and initial stress: The effect of having a strong initial syllable is stronger when the two words are less phonologically similar. In sum, the interaction between metrical structure and segmental structure is not limited to initial segmental overlap; instead, it spans the whole word. Thus, by calculating phonological distance in addition to manipulating initial segmental overlap, we provided further evidence for the joint planning of metrical structure and segmental structure.

Nonetheless, unlike what we found in the initial segmental overlap models, there was no interaction effect between metrical overlap and phonological distance. One possibility is that initial segments play a more important role than other segments, and taking into account all the segments masks the interaction effect. Another potential explanation is the unbalanced distribution of word pairs across different phonological distance values. Because phonological distance is not a controlled variable in the current study, the number of word pairs across different phonological distances vary. For instance, while there are 45 word pairs that have a phonological distance of 5, only two word pairs have a phonological distance of 7, and only four

word pairs have a phonological distance of 1. This unbalanced distribution may mask some interaction effects between phonological distance and other variables, especially because the two ends of the continuum (1 and 7) are not well-represented in our data. Therefore, future research should consider manipulating phonological distance between words to further investigate whether phonological distance has an effect on speech planning difficulty.

Comparison Between the Two Sets of Models

For both mean word pair duration and number of repetitions, the initial segmental overlap models outperformed the phonological distance models, as evidenced by higher adjusted R^2 values (0.031 vs. 0.024 for mean word pair duration, and 0.034 vs. 0.032 for number of repetitions). Thus, accounting for segmental overlap in other positions does not significantly improve the model's predictive power for speech planning. Therefore, the current analysis provided further support for the importance of the initial segment in lexical access, resonating with previous work showing that initial segments are important for word processing (Marslen-Wilson & Welsh, 1978). Moreover, previous research has shown that phonemes are activated sequentially, from left to right (Sevold & Dell, 1994). Therefore, using the initial segments during word identification is efficient: As soon as a segment in a candidate word does not match with the intended word, the candidate word can be depressed to a lower activation level, shrinking the number of highly activated candidate words.

Nevertheless, it needs to be noted that our results only suggest that segmental overlap in other positions may not be as influential as initial segmental overlap, but do not suggest that segmental overlap in other positions does not matter. In fact, many previous studies have shown qualitatively similar effects of initial segments and final segments on word processing. For example, Allopenna and colleagues (1998) showed that words that do not share initial segments

but rhyme with the heard word are activated. Damian and Dumay (2007) observed a shortened speech onset latency when the distractor word and the target word shared either the first syllable or the last syllable. Brown and McNeil (1966) found that when people experience the Tip-of-the-Tongue, even though it is more often that they can retrieve the initial segment of the word, they sometimes can also retrieve the final segment. Because we did not control for whether the words in a word pair share final segments, our data did not allow us to assess whether initial segmental overlap and final segmental overlap have similar effects on word pair repetition. Future research is needed to better understand how the segment position may affect the segmental overlap effect, which will shed light on the nature of the phonological encoding process.

In sum, our results are in line with the account that metrical structure and segmental structure are jointly represented, consistent with the conclusions drawn in Yiu and Watson (2015), Myers and Watson (2019), and Myers and Watson (2021). In other words, each syllable would have two versions: a stressed version and an unstressed version. In lexical representation, one of the two versions of a syllable is specified for each syllable position. That is, when we access the word “arrow,” we know that it consists of two syllables: the stressed version of /ε/ and the unstressed version of /rou/. Therefore, our results are inconsistent with the WEAVER model’s proposal about the planning of metrical structure and segmental structure which states that there are two steps in the planning of metrical and segmental structures: The first step is the parallel retrieval of the metrical structure and the segmental structure and the second step is the association of the retrieved structures (Roelofs, 1997). Instead, our results suggest that the metrical structure and the segmental structure are retrieved together, and thus, no association is needed after the retrieval.

We did not find any facilitatory effects of initial segmental overlap nor metrical overlap, contradicting previous findings (Damian & Dumay, 2007; Tilsen, 2011). One possible explanation for the contradictory results lies in the methodology. In Damian and Dumay (2007), participants produced spontaneous speech to describe a picture, whereas in our experiment, participants repeated scripted word pairs. As shown in Allopenna and colleagues' (1998) work, being presented with a phonologically related word, either in the first syllable position or the final syllable position, would activate the target noun word in the intended noun phrase and thus speed up speech planning in Damian and Dumay's (2007) experiment. However, in the current experiment, the two words were already activated well before the speech production started: Participants read the word pair silently for 4 seconds before starting to repeat the word pair out loud. Although Tilsen (2011) also used a repetition task, there are two important differences between our experiments that may contribute to the different results. First, sequences of four trisyllabic nonwords were used in Tilsen's (2011) experiment, while pairs of two real words were used in our experiment, meaning that there was lexical access in our experiment but not in Tilsen (2011). Thus, our different results suggest that the interference effect of metrical overlap we observed may be specific to lexical processing but not to nonword processing, allowing us to use our results to draw inferences on lexical representation. Second, in Tilsen (2011), the stressed vowels were indicated by "e" (e.g., "meetida" has a Strong-weak-weak stress pattern), meaning that there was a cue specifically associated with the stress pattern of the nonword sequences. On the contrary, participants in our experiment did not get any cue on the stress pattern of the word pairs. Therefore, the experiment in Tilsen (2011) might guide participants to pay more attention to the metrical structure of the nonword sequence. It is possible that this greater attention to the metrical structure, together with the fact that there was no lexical access

involved in Tilsen's (2011) experiment, led to the observed facilitatory effect of metrical structure.

In addition to the previously discussed implications for speech planning and lexical representation theories, the current study also offered insight into how metrical sensitivity may facilitate word processing. Because metrical structure and segmental structure are closely bound in lexical representation, better recognition of the metrical structure may make the recognition of the segments easier, accelerating the overall word processing. Although our current individual differences analysis does not suggest how music processing skill may affect word pair repetition, many previous studies have found that metrical sensitivity predicts reading comprehension in children, after controlling for other important factors (e.g., age, phonological awareness, etc.). For example, it was found that for children aged from 5 to 7, receptive sensitivity to speech rhythm and non-speech rhythm each predicted a significant amount of unique variance in word reading ability over and above age, vocabulary, phonological awareness, and short-term memory (Holliman et al., 2010). Similarly, for children aged from 4 to 7, better sensitivity to metrical stress reversal in words predicted better literacy, phonological awareness, and spelling ability (Wood, 2006). Moreover, the influence of rhythm on language development is present even before children learn how to read. Starting in infancy, language-specific rhythm and stress are well recognized and used: Infants as young as 4 days old already use linguistic rhythm and stress to differentiate their native language (in this case, French) from other languages (Mehler et al., 1988). Language-specific rhythm and stress cues may later be used for spoken word processing (Cutler & Mehler, 1993). For instance, recognizing that the majority of English words start with a stressed syllable allows infants to use rhythm to identify word boundaries: The onset of a new word is usually a stressed syllable.

Indeed, children are getting a lot of cues to the metrical structure in language in daily life. One source is child-directed speech (CDS). One important feature of CDS is that it has a larger fundamental frequency (F0) range than adult-directed speech (ADS; Vosoughi & Roy, 2012). F0 is an important cue for signaling word stress, along with longer vowel durations and higher intensity. Thus, by increasing the F0 range, the difference between stressed syllables and unstressed syllables would be exaggerated in CDS. That is, the metrical pattern is more pronounced in CDS than in ADS, supporting children's speech processing. Another source is children's literature, especially the metrically regular and rhyming language that is ubiquitous in children's literature (Beck & Konieczny, 2021). Taking previous work together, it is clear that children are able to pick up metrical cues and use them in language processing and development. Therefore, future work should explore how rhythmic and metrical training can be effectively incorporated into literacy development programs.

An Alternative Explanation of the Current Results

Although the joint representation of metrical structure and segmental structure explains well why the speech production was more challenging when there is metrical overlap or segmental overlap, this is not the only explanation. An alternative explanation is that the challenges that segmental overlap and metrical overlap cause come from two different sources: segmental overlap word pairs are difficult to produce due to motor constraint, and metrical overlap word pairs are difficult to produce due to rhythmic constraint.

Motor Constraint. We define motor constraint as the limit on how fast our muscles can move to carry out two adjacent similar movements. Some insights into motor constraint are provided by previous finger-tapping studies. It was observed that a healthy individual can tap about 50-60 times per 10 s using the index finger of their dominant hand (Cousins et al., 1998;

Da Silva et al., 2012; Schmidt et al., 2000). Finger-tapping rate has also been studied in the context of sensorimotor synchronization which is defined as the coordination of an action with a rhythmic stimulus sequence (e.g., auditory clicks, visual flashes of light, etc.). Bartlett and Bartlett (1959) asked participants to make a single tap that coincided with any click in an isochronous click sequence. They found that all three participants failed to synchronize with the click sequence when the inter-onset interval (IOI) was 125 ms, suggesting that this IOI is shorter than the fastest IOI their finger tapping can synchronize to. Together, these findings suggest that there exists an upper limit on the finger tapping rate.

Similarly, muscle movement is involved in speech production: Our lips, tongue, teeth, glottis, lungs, and many other organs are actively engaged when we produce sounds. Therefore, the constraint on how fast these muscles can move would also result in a constraint on how short the interval between two adjacent sounds can be. This interval would also be affected by the specific sounds involved. For example, two similar sounds occurring close to each other is harder to produce than two dissimilar sounds occurring close to each other. In fact, there is a phonological change called dissimilation that is specifically responsible for making two similar sounds in a word less similar. Given that our experiment requires participants to repeat short word pairs at a fast speech rate, when there is segmental overlap, the same sound needs to be produced temporally close to each other. Therefore, these word pairs including segmental overlap would require a longer duration than word pairs without segmental overlap, simply because some delay is needed to produce the similar sounds. Because this constraint is not limited to initial segmental overlap, it accounts well for our findings that segmental overlap in any position made the production more challenging and that the more segmental overlaps there were in the word pair, the more challenging the production was.

Rhythmic Constraint. Inspired by Cummins and Port (1998), we regard rhythmic constraint as the constraint that speech rhythm imposes on the timing of lower-level speech units. In other words, speech rhythm is a higher-level dynamic within which the timing of the lower-level smaller units of speech production are constrained. An analogy they used to explain this idea is that speech rhythm is like the steering wheel of a car and the smaller speech units are like the four wheels of the car. It is hard for drivers to control four wheels because they can each point in any direction (i.e., there are four degrees of freedom). However, with a steering wheel that controls all four wheels together, there is only one degree of freedom, making the task for drivers much easier. Similarly, in speech production, smaller speech units can in principle be produced at any time, making it challenging to organize the timing of every individual speech unit. However, with speech rhythm functioning as a higher-level coordinating mechanism that organizes and constrains the timing of these lower-level speech units, the degree of freedom is reduced, making the temporal organization of speech production much easier.

Cummins and Port (1998) also showed experimental results that suggest that a lower-level dynamic in speech production exists, a higher-level dynamic in speech production exists, and this higher-level dynamic lies in speech rhythm. During their speech cycling experiments, auditory high and low tones were played. They defined the interval between two adjacent high tones (i.e., from the onset of a high tone to the onset of the next high tone) as a cycle that has a phase value of 1. The duration of the high tone was fixed while the duration of the low tone varied, such that the target phase value of the low tone duration was randomly selected from a uniform distribution between 0.3 and 0.7. These tone sequences were played and participants were asked to repeat short phrases (e.g., “big for a duck”) in coordination with the tone sequence. Participants continued repeating the phrase after the tone sequence ended. About half of the

repetitions were produced while the tone sequence was present, and about half of the repetitions were produced while the tone sequence was absent. They measured the phase value of the interval from the onset of the second stressed syllable in one repetition to the onset of the first stressed syllable in the next repetition (e.g., from the onset of “duck” to the onset of following “big”) and compared the observed phase value to the target phase value.

They found that while in principle participants can produce any phase values between 0.3 and 0.7, in actuality they produced two or three phases reliably at values close to $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$. This within- and between-subject convergence of preferred phase suggests that a higher-level dynamic exists in speech production. The rhythmic nature of the experiment further suggests that this higher-level dynamic is about speech rhythm. Moreover, the phrase was produced with different prosodic structures depending on the preferred phase. For phase values close to $\frac{1}{3}$ and $\frac{1}{2}$, the phrase is divided into two feet: [big for a] [duck]. The difference between these two phase values is that, when the phase value is close to $\frac{1}{3}$, a silent beat is inserted, such that the prosodic structure is [big for a] [duck] [], whereas when the phase value is close to $\frac{1}{2}$, the silent beat is not needed, such that the prosodic structure is simply [big for a] [duck]. The production for phase values close to $\frac{2}{3}$ is more different from the above-mentioned patterns: The phrase is divided into three feet: [big] [for a] [duck]. Therefore, a phase value specified in the speech rhythm (i.e., the higher-level dynamic) coordinates the metrical feet (i.e., the lower-level dynamic), supporting the existence of rhythmic constraints on stress timing in English.

Given that our experiment also has a repetitive nature and that there exists a fixed metrical structure within each word pair, the rhythmic constraint proposed by Cummins and Port (1998) should apply. Thus, it is possible that word pairs without metrical overlap have fewer

phases than word pairs with metrical overlap. One possible way for this to happen is for Strong-weak sequences to take roughly the same time as Strong-weak-weak sequences. Indeed, Fant and Kruckenberg (1996) found that, during reading aloud in Swedish, the average duration of interstress interval was about 540 ms, the average duration of stressed syllables was 270 ms, and the average duration of unstressed syllables was 132 ms. According to these durations, a Strong-weak sequence, if followed by a stressed syllable, should take about 540 ms because this is one interstress interval. A Strong-weak-weak sequence, if followed by a stressed syllable, should also only take about 540 ms for two reasons. First, at the surface level, the total duration of a stressed syllable and two unstressed syllable is about 534 ms ($270 + 132 + 132 = 534$). Second, given that this sequence is followed by another stressed syllable, the duration of this sequence is also one interstress interval. In other words, the duration of a Strong-weak-weak sequence can be the same as the duration of a Strong-weak sequence. Therefore, each word pair with metrical overlap has two cycles: Strong-weak Strong-weak word pairs have [Strong-weak] [Strong-weak] and weak-Strong weak-Strong word pairs have [weak-Strong] [weak-Strong]. On the contrary, each word pair without metrical overlap has less than two cycles. Strong-weak weak-Strong word pairs have [Strong-weak-weak] [Strong], where [Strong-weak-weak] is one cycle and [Strong] is shorter than one cycle because it has a shorter duration than [Strong-weak]. For weak-Strong-Strong-weak pairs, the [Strong-weak] part at the end of a repetition and the [weak] part at the start of the next repetition can form a [Strong-weak-weak] cycle. Therefore, on average, a weak-Strong Strong-weak word pair would also have fewer than two cycles. Hence, word pairs without metrical overlap would be easier to produce than word pairs with metrical overlap. This hypothesis can be tested by measuring the duration of each syllable and comparing whether [Strong-weak-weak] syllable sequences have similar duration to [Strong-weak] syllable

sequences, regardless of whether the sequence is within a repetition or spans across repetitions.

Individual Differences and Speech Planning

Individual Differences in WMC

Our hypothesis that there would be an interaction effect between WMC and metrical overlap and an interaction effect between WMC and initial segmental overlap was not supported. We would need more data to determine whether this lack of effect is due to a weak statistical power.

However, we found a main effect of WMC on all three dependent variables: People with better WMC had longer mean word pair duration and fewer repetitions but fewer errors than people with worse WMC. One possible explanation is that people with better WMC were indeed able to plan the whole word pair at once. Thus, they did not need to remap the speech plans, resulting in fewer errors. At the same time, however, planning two words required more time than planning only one word, leading to a longer mean word pair duration and fewer repetitions. Nonetheless, one piece of evidence against this explanation is Arnold and Nozari's experiment exploring the causal relationship between working memory and message planning (Arnold & Nozari, 2017). They found that while participants who got sham Transcranial Direct Current Stimulation (tDCS) required longer speech onset latency to produce more coherent discourse, participants who got anodal tDCS to stimulate left prefrontal cortex (a brain region found to be associated with working memory) were able to produce similar coherent discourse without a longer latency. That is, enhancing working memory by activating the associated brain region enables better message-level speech planning without a time cost. However, message-level planning and phonological planning are very different, and thus the role of WMC in these two processes may also be very different. For example, previous picture-word interference

experiments have shown that verbal working memory load reduced the scope of phonological planning, but not semantic planning (Klaus et al., 2017).

Individual Differences in Music Perception Skill

Our hypothesis that there would be an interaction effect between music perception skill and metrical overlap was not supported. More data are needed to determine if this lack of effect is due to the small sample size. However, we found a main effect of music perception skill on mean word pair duration and number of errors: People with better music perception skill had shorter mean word pair duration but produced more errors.

The opposite effects of WMC and music perception skill we observed can be explained by the distribution of WMC and music perception skill in the participants included in the current analysis. Out of the 10 participants, only one of them got high scores for both WMC and music perception skill, only one of them got low scores for both WMC and music perception skill, four of them got a high WMC score but low music perception score, and four of them got a high music perception score but a low WMC score. In other words, the majority of the participants is good at one individual difference measure and bad at the other. Therefore, the effect we found of having a high WMC score is likely the effect of having a high WMC score and a low music perception score; similarly, the effect we found of having a high music perception score is likely the effect of having a high music perception score and a low WMC score. Fortunately, an examination of the 58 participants who will be included in the future analysis showed that there is no overall correlation between an individual's WMC score and music perception score ($r = 0.051, p = 0.7$). Therefore, we anticipate being able to uncover more meaningful effects of individual differences measures in the future analysis.

Conclusion

Using a tongue-twister word pair repetition paradigm, the current study explored whether metrical structure and segmental structure are independently or jointly planned during English speech production. We found that speech production was more difficult when there was either metrical overlap or segmental overlap. Furthermore, speech production was even more challenging when there were both metrical and segmental overlaps. Thus, replicating and extending Myers and Watson's (2021) findings, our results favor the account that metrical structure and segmental structure are jointly planned during phonological encoding. Our results also suggest that metrical structure is bound to segmental structure in lexical representation. That is, each syllable has two versions: a stressed version and an unstressed version. This finding suggests that better metrical sensitivity may facilitate segmental recognition, enhancing word processing skills. Therefore, future work should be done to explore how metrical training may be effectively integrated into literacy development programs. In addition, we also found that speech production was easier when the first syllable of the word pair was stressed and that segmental overlap, not only in the initial position, but in any position, affected speech planning difficulty. These results further suggest that speech planning is complicated and can be influenced by many factors. In future analysis, our data on individual differences in WMC and music perception skills will allow us to reveal how these individual differences may affect phonological encoding and speech production, providing insights into the non-linguistic factors that affect speech production difficulty.

APPENDIX

Table A1*Word Pairs with Initial Segmental Overlap*

Metrical Overlap		No Metrical Overlap	
Initial Stress: S SwSw	Initial Stress: w wSwS	Initial Stress: S SwwS	Initial Stress: w 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

Table A2*Word Pairs Without Initial Segmental Overlap*

Metrical Overlap		No Metrical Overlap	
Initial Stress: S SwSw	Initial Stress: w wSwS	Initial Stress: S SwwS	Initial Stress: w wSSw
candy arrow	cavort array	candle arrive	canoe error
passive ballot	possess ballet	pastor balloon	pastel ballad
relish beaker	relate become	relic because	relax beagle
seeker candle	secrete canoe	secret cavort	secure candy
purchase carrot	persist career	purple caress	pursuit carriage
missive college	misstep collage	mister collide	mistake column
despot concept	design concern	desert conceal	dessert concert
office corner	afoot correct	offer corrupt	offend coral
error desert	arrive dessert	arrow design	array despot
info dismal	infer dismay	infant dismiss	inform distant
parent exit	parole exam	parrot excel	parade extra
carriage infant	caress inform	carrot infer	career info
concert mister	conceal mistake	concept misstep	concern missive
extra offer	excel offend	exit afoot	exam office
distant parrot	dismiss parade	dismal parole	dismay parent
coral pastor	corrupt pastel	corner possess	correct passive
reason purple	recede pursuit	recent persist	recite purchase
column recent	collide recite	college recede	collage reason
beagle relic	because relax	beaker relate	become relish
ballad secret	balloon secure	ballot secrete	ballet seeker

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>
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. <https://doi.org/10.1006/jmla.1997.2558>
- Arnold, J. E., & Nozari, N. (2017). The effects of utterance timing and stimulation of left prefrontal cortex on the production of referential expressions. *Cognition*, *160*, 127–144. <https://doi.org/10.1016/j.cognition.2016.12.008>
- Bartlett, N. R., & Bartlett, S. C. (1959). Synchronization of a motor response with an anticipated sensory event. *Psychological Review*, *66*(4), 203–218. <https://doi.org/10.1037/h0046490>
- Beck, J., & Konieczny, L. (2021). Rhythmic subvocalization: An eye-tracking study on silent poetry reading. *Journal of Eye Movement Research*, *13*(3), 10.16910/jemr.13.3.5. <https://doi.org/10.16910/jemr.13.3.5>
- Besson, M., & Schön, D. (2001). Comparison between Language and Music. *Annals of the New York Academy of Sciences*, *930*(1), 232–258. <https://doi.org/10.1111/j.1749-6632.2001.tb05736.x>
- Bishop, J. (2021). Exploring the Similarity Between Implicit and Explicit Prosody: Prosodic Phrasing and Individual Differences. *Language and Speech*, *64*(4), 873–899. <https://doi.org/10.1177/0023830920972732>
- Bishop, J., & Intlekofer, D. (2020). Lower Working Memory Capacity is Associated with Shorter Prosodic Phrases: Implications for Speech Production Planning. *Speech Prosody 2020*, 191–195. <https://doi.org/10.21437/SpeechProsody.2020-39>
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. [https://doi.org/10.1016/0010-0285\(86\)90004-6](https://doi.org/10.1016/0010-0285(86)90004-6)
- Boersma, Paul & Weenink, David (2023). Praat: doing phonetics by computer. Version 6.3.17, retrieved 28 Spetember 2023 from <http://www.praat.org/>
- Brown, R., & McNeill, D. (1966). The “Tip of the Tongue” Phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 325–337. [https://doi.org/10.1016/S0022-5371\(66\)80040-3](https://doi.org/10.1016/S0022-5371(66)80040-3)
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. <https://doi.org/10.3758/BF03196772>
- Cousins, M. S., Corrow, C., Finn, M., & Salamone, J. D. (1998). Temporal Measures of Human Finger Tapping: Effects of Age. *Pharmacology Biochemistry and Behavior*, *59*(2), 445–449. [https://doi.org/10.1016/S0091-3057\(97\)00443-7](https://doi.org/10.1016/S0091-3057(97)00443-7)
- Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of*

- Phonetics*, 26(2), 145–171. <https://doi.org/10.1006/jpho.1998.0070>
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2(3–4), 133–142. [https://doi.org/10.1016/0885-2308\(87\)90004-0](https://doi.org/10.1016/0885-2308(87)90004-0)
- Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, 21(1–2), 103–108. [https://doi.org/10.1016/S0095-4470\(19\)31323-3](https://doi.org/10.1016/S0095-4470(19)31323-3)
- Da Silva, F. N., Irani, F., Richard, J., Brensinger, C. M., Bilker, W. B., Gur, R. E., & Gur, R. C. (2012). More than just tapping: Index finger-tapping measures procedural learning in schizophrenia. *Schizophrenia Research*, 137(1–3), 234–240. <https://doi.org/10.1016/j.schres.2012.01.018>
- Damian, M., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, 57(2), 195–209. <https://doi.org/10.1016/j.jml.2006.11.001>
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Fant, G., & Kruckenberg, A. (1996). On the quantal nature of speech timing. *4th International Conference on Spoken Language Processing (ICSLP 1996)*, 2044–2047. <https://doi.org/10.21437/ICSLP.1996-518>
- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, 37(1), 1–9. <https://doi.org/10.3758/MC.37.1.1>
- Féry, C. (1998). German Word Stress in Optimality Theory. *The Journal of Comparative Germanic Linguistics*, 2, 101–142. <http://www.jstor.org/stable/23738532>
- FindingFive Team (2023). FindingFive: An online platform for creating, running, and managing your experiments. <https://www.findingfive.com>
- Gordon, R. L., Magne, C. L., & Large, E. W. (2011). EEG Correlates of Song Prosody: A New Look at the Relationship between Linguistic and Musical Rhythm. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00352>
- Heffner, C. C., & Slevc, L. R. (2015). Prosodic Structure as a Parallel to Musical Structure. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01962>
- Holliman, A. J., Wood, C., & Sheehy, K. (2010). The contribution of sensitivity to speech rhythm and non-speech rhythm to early reading development. *Educational Psychology*, 30(3), 247–267. <https://doi.org/10.1080/01443410903560922>
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149. <https://doi.org/10.1037/0033-295X.99.1.122>
- Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3(3), 129–140. [https://doi.org/10.1016/S0095-4470\(19\)31360-9](https://doi.org/10.1016/S0095-4470(19)31360-9)
- Klaus, J., Mädebach, A., Oppermann, F., & Jescheniak, J. D. (2017). Planning sentences while doing other things at the same time: Effects of concurrent verbal and visuospatial working memory load. *Quarterly Journal of Experimental Psychology*, 70(4), 811–831. <https://doi.org/10.1080/17470218.2016.1167926>
- Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26(1), 1–37. https://doi.org/10.1207/s15516709cog2601_1
- 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>
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics. Doklady*, 10(8), 707–710. <https://ci.nii.ac.jp/naid/10020212767>
- Marie, C., Magne, C., & Besson, M. (2011). Musicians and the Metric Structure of Words. *Journal of Cognitive Neuroscience*, 23(2), 294–305.
<https://doi.org/10.1162/jocn.2010.21413>
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1), 29–63.
[https://doi.org/10.1016/0010-0285\(78\)90018-X](https://doi.org/10.1016/0010-0285(78)90018-X)
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, 498–502. <https://doi.org/10.21437/Interspeech.2017-1386>
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, B., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143–178.
[https://doi.org/10.1016/0010-0277\(88\)90035-2](https://doi.org/10.1016/0010-0277(88)90035-2)
- Meyer, A. S., & van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, 7(2), 314–319. <https://doi.org/10.3758/BF03212987>
- Moss, C., Kwabi, S., Ardoin, S. P., & Binder, K. S. (2024). Eye movements and reading comprehension performance: Examining the relationships among test format, working memory capacity and reading comprehension. *Reading and Writing*, 37(3), 703–729.
<https://doi.org/10.1007/s11145-023-10428-0>
- Myers, B. R., & Watson, D. G. (2019). Paying the meter: Effect of metrical similarity on word lengthening. *Psychonomic Bulletin & Review*, 26(6), 1941–1947.
<https://doi.org/10.3758/s13423-019-01635-4>
- Myers, B. R., & Watson, D. G. (2021). Evidence of Absence: Abstract Metrical Structure in Speech Planning. *Cognitive Science*, 45(8), e13017. <https://doi.org/10.1111/cogs.13017>
- Peretz, I., Champod, A., & Hyde, K. (2003). Varieties of Musical Disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, 999, 58–75. <https://doi.org/10.1196/annals.1284.006>
- R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- 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](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., Weber, R. J., Hazelett, W. M., & Hindorff, V. (1986). The parameter remapping effect in human performance: Evidence from tongue twisters and finger fumlbers. *Journal of Memory and Language*, 25, 710–725. [https://doi.org/10.1016/0749-596X\(86\)90045-8](https://doi.org/10.1016/0749-596X(86)90045-8)
- Schmidt, S. L., Oliveira, R. M., Krahe, T. E., & Filgueiras, C. C. (2000). The effects of hand preference and gender on finger tapping performance asymmetry by the use of an infra-

- red light measurement device. *Neuropsychologia*, 38(5), 529–534.
[https://doi.org/10.1016/s0028-3932\(99\)00120-7](https://doi.org/10.1016/s0028-3932(99)00120-7)
- Schneider, K., Dogil, G., & Möbius, B. (2011). Reaction time and decision difficulty in the perception of intonation. *Interspeech 2011*, 2221–2224.
<https://doi.org/10.21437/Interspeech.2011-581>
- Sevold, 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](https://doi.org/10.1016/0010-0277(94)90067-1)
- Slevc, L. R., & Miyake, A. (2006). Individual Differences in Second-Language Proficiency: Does Musical Ability Matter? *Psychological Science*, 17(8), 675–681.
<https://doi.org/10.1111/j.1467-9280.2006.01765.x>
- Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2), 374–381. <https://doi.org/10.3758/16.2.374>
- Sun, M., Xing, W., Yu, W., Slevc, L. R., & Li, W. (2024). ERP evidence for cross-domain prosodic priming from music to speech. *Brain and Language*, 254, 105439.
<https://doi.org/10.1016/j.bandl.2024.105439>
- Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136(1), 64–81. <https://doi.org/10.1037/0096-3445.136.1.64>
- Tilsen, S. (2011). Metrical regularity facilitates speech planning and production. *Laboratory Phonology*, 2(1). <https://doi.org/10.1515/labphon.2011.006>
- Tooley, K. M., Konopka, A. E., & Watson, D. G. (2014). Can Intonational Phrase Structure be Primed (like Syntactic Structure)? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 348–363. <https://doi.org/10.1037/a0034900>
- Tooley, K. M., Konopka, A. E., & Watson, D. G. (2018). Assessing priming for prosodic representations: Speaking rate, intonational phrase boundaries, and pitch accenting. *Memory & Cognition*, 46(4), 625–641. <https://doi.org/10.3758/s13421-018-0789-5>
- Tsukahara, Jason S. (2022). englelab: An R package for processing complex-span and attention control tasks downloaded from the EngleLab (1.1.0). Zenodo.
<https://doi.org/10.5281/zenodo.6987145>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505.
<https://doi.org/10.3758/BF03192720>
- Vosoughi, S., & Roy, D. (2012). A longitudinal study of prosodic exaggeration in child-directed speech. *Speech Prosody 2012*, 194–197. <https://doi.org/10.21437/SpeechProsody.2012-50>
- Vuvan, D. T., Paquette, S., Mignault Goulet, G., Royal, I., Felezeu, M., & Peretz, I. (2018). The Montreal Protocol for Identification of Amusia. *Behavior Research Methods*, 50(2), 662–672. <https://doi.org/10.3758/s13428-017-0892-8>
- 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: Studies in Honor of Janet Dean Fodor* (pp. 85–98). Springer International Publishing. https://doi.org/10.1007/978-3-319-12961-7_5
- Wilshire, C. E. (1999). The “Tongue Twister” Paradigm as a Technique for Studying Phonological Encoding. *Language and Speech*, 42(1), 57–82.
<https://doi.org/10.1177/00238309990420010301>

- Wood, C. (2006). Metrical stress sensitivity in young children and its relationship to phonological awareness and reading. *Journal of Research in Reading*, 29(3), 270–287. <https://doi.org/10.1111/j.1467-9817.2006.00308.x>
- 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>